TP 7360E

APPLICATION OF VOICE TECHNOLOGY TO TRANSPORTATION OF DISABLED PERSONS

Prepared for the

Transportation Development Centre

Transport Canada

by TransVision Consultants Ltd.

January 1986

Library and Information Centre / Centre de Documentation (GSLA)
RECEIVED · REÇU LE
JUL 17 1986
Transport Canada / Transports Canada

The contents of this report reflect the views of the author and not necessarily the official views of the Transportation Development Centre

Un sommaire en français de ce rapport est inclus avant la table des matières

REPORT DOCUMENTATION FORM

1. Transport Canada Report No.	2. TDC Project No	3. Recipient's Catalogue No.
TP 7360E	6352	

4. Title and Subtitle

Application of Voice Technology to Transportation of Disabled Persons

5. Report Date
January 1986

6. Performing Organization Report No.

7. Author(s)
Tom Geehan

8. Transport Canada File No.
D1450-103-60-4

9. Performing Organization Name and Address
TransVision Consultants Ltd.
3230 Chaucer Avenue
North Vancouver, B.C. V7K 2C3

10. DSS File No.
15SD.T8200-3-3575-03

11. DSS or Transport Canada Contract No.
OSD84-00260

12. Sponsoring Agency Name and Address
Transportation Development Centre (TDC)
200 Dorchester Blvd. West
Suite 601, West Tower
Montreal, Quebec H2Z 1X4

13. Type of Report and Period Covered

14. Sponsoring Agency Code
TDC 6352

15. Supplementary Notes

16. TDC Project Officer
Barbara Smith/Ling Suen

17. Abstract

Computer-based speech is an increasingly important area in information processing, since it utilizes the form of communication most natural to people -- speech. This report examines three distinct areas: voice recognition, voice synthesis, and voice compression. Individual speech techniques are reviewed, current applications for speech examined, specific applications developed for improving the mobility of the elderly and disabled, and directions offered for planning and implementing a program of research and development. Computer-based speech is appropriate for both control and information functions in the transportation of the elderly and disabled. The report contains detailed comparison charts featuring over a hundred speech products from about fifty different vendors.

18. Key Words	19. Distribution Statement
Computers, Voice Technology, Disabled, Physically Handicapped, Transportation	Limited number of copies available from the Transportation Development Centre

20. Security Classification (of this report)	21. Security Classification (of this page)	22. No. of Pages	23. Price
Unclassified	Unclassified	86	Nil

TDC CDT 79-005

FORMULE DE DOCUMENTATION POUR RAPPORT

1. N° du rapport Transports Canada	2. N° de l'étude — CDT	3. N° de catalogue du destinataire
TP 7360E	6352	

4. Titre et sous-titre	5. Date du rapport
Application of Voice Technology to Transportation of Disabled Persons	Janvier 1986
	6. N° du rapport de l'organisme exécutant

7. Auteur(s)	8. N° de dossier Transports Canada
Tom Geehan	D1450-103-60-4

9. Nom et adresse de l'organisme exécutant	10. N° de dossier — ASC
TransVision Consultants Ltd. 3230 Chaucer Avenue North Vancouver, B.C. V7K 2C3	15SD.T8200-3-3575-03
	11. N° de contrat — ASC ou Transports Canada OSD84-00260

12. Nom et adresse de l'organisme parrain	13. Genre de rapport et période visée
Centre de développement des transports (CDT) 200 ouest, boul. Dorchester Suite 601, tour ouest Montréal, Québec H2Z 1X4	14. Code de l'organisme parrain TDC 6352

15. Remarques additionnelles	16. Agent de projet — CDT
	Barbara Smith/Ling Suen

17. Résumé

Le traitement automatique de la parole occupe une place de plus en plus grande dans le domaine de traitement de l'information, étant donné qu'il est fondé sur le mode de communication le plus naturel chez l'être humain : la parole. Les chercheurs se sont penchés sur les trois grands domaines que sont la reconnaissance de la parole, la synthèse de la parole et la compression de la parole. Ils ont fait un tour d'horizon des différentes techniques existantes, des applications courantes et de celles qui amélioreraient le sort des handicapés et des personnes âgées. Ils ont aussi formulé des suggestions relatives à l'élaboration et la mise en oeuvre d'un programme de recherche et de développement. Les systèmes en question peuvent être utiles tant pour la commande orale d'appareils et instruments que pour fournir de l'information. Le rapport présente des tableaux comparatifs concernant plus d'une centaine de produits proposés par une cinquantaine de fabricants.

18. Mots-clés	19. Diffusion
ordinateurs, traitement automatique de la parole, invalides, handicapés, transport.	Le CDT dispose d'un nombre limité d'exemplaires.

20. Classification de sécurité (de ce rapport)	21. Classification de sécurité (de cette page)	22. Nombre de pages	23. Prix
Non classifié	Non classifiée	86	Aucun

SUMMARY

The Transportation Development Centre, in recognition of a common concern for increasing independence by individuals with disabilities, commissioned this study into the application of voice technology to the transportation problems of the disabled. This study looks at voice technology as one of several related technologies encompassing computers, robotics, artificial intelligence and natural language processors that can be used to develop a positive change in the lifestyles of elderly and disabled persons.

Computer-based speech is an increasingly important area in information processing and control systems, since it utilizes the form of communication most natural to people -- speech. This report examines three distinct areas: voice recognition, voice synthesis, and voice compression. Individual speech techniques are reviewed, current applications for speech examined, specific applications developed for improving the mobility of the elderly and disabled, and directions offered for planning and implementing a research and development program.

The need to have control of our environment and control of our lives, in order to reduce dependence, encompasses two types of control. First, it is access to and control of objects in the individuals **immediate** physical surroundings, often by using assistive devices. Second, it is access to and control of items **outside** of the individual's immediate surroundings through the applications of communications technology.

The first major application of voice technology is in access to and control of objects in the individual's immediate physical surroundings by using other assistive devices controlled by the voice units. The second major

application, the one of primary interest in this report, is access to and control of items outside of the individual's immediate surroundings through the application of telecommunication systems. A brief review of the requirements for improved transportation services for the elderly and disabled reveals a preponderance of requirements for information.

Various applications of voice technology to transportation for the elderly and disabled were identified in this study, including: telephone voice response systems, personal transportation terminal directional systems, voice-controlled wheelchairs, and talking transportation assistive devices. The study also identified a need for more research on the acceptability of the technology to the end user.

Clearly, the next step would be to evaluate the technical, operational, and economic feasibility of the potential applications -- either through further intensive research or through selected demonstration projects. Research statements are provided in the study for each proposed application in order to provide directions for this research effort.

In conclusion, the study recommends that the Transportation Development Centre initiate a research program oriented towards the application of voice technology to the transportation needs of the elderly and disabled. Several potential applications have been identified and there are undoubtedly many more. The problem of acceptance by individual users will have to be addressed. The general resistance to new technology in any application must be overcome as well as specific problems. User acceptance is affected by the anticipated benefits and methods of presentation and implementation, as well as device reliability, recognition robustness, and naturalness of synthesized speech.

Technological advances are expected to address these latter problems so that voice technology will become increasingly comfortable to use. A research program is necessary to address some of the issues raised by the other problems. Finally, applications should be based on documented and demonstrated performance on standard feasibility tests and a thorough understanding of the effort required to implement the application, especially when voice will significantly modify the way users have accessed transportation systems and services in the past.

SOMMAIRE

Conscient de la ferme volonté des personnes handicapées d'accroître leur autonomie, le Centre de développement des transports a lancé la présente étude portant sur les systèmes de traitement automatique de la parole comme moyen susceptible de réaliser cet objectif. Pour les chercheurs, cette technique n'a été qu'une des différentes technologies envisagées, à savoir l'informatique, la robotique, l'intelligence artificielle et les processeurs de langues naturelles, toutes pouvant servir à aider les personnes âgées et les handicapés à améliorer leur sort.

Les systèmes de traitement automatique de la parole occupent une place de plus en plus grande dans le domaine du traitement de l'information, étant donné qu'ils sont fondés sur le mode de communication le plus naturel chez l'être humain : la parole. Les chercheurs se sont penchés sur les trois grands domaines que sont la reconnaissance de la parole, la synthèse de la parole et la compression de la parole. Ils ont fait un tour d'horizon des différentes techniques existantes, des applications courantes et de celles qui amélioreraient le sort des personnes âgées et des handicapés désireux de se déplacer, et ils ont formulé des suggestions relatives à l'élaboration et la mise en oeuvre d'un programme de recherche et de développement.

Pour dominer notre environnement et diriger notre vie comme nous l'entendons, bref, pour accroître notre autonomie, chacun de nous doit être capable de maîtriser deux grandes catégories d'obstacles physiques : ceux directement à notre portée et ceux en dehors de notre portée. La maîtrise des premiers peut être obtenue par des dispositifs qui sont autant d'aides et celle des deuxièmes par le moyen de la communication.

SOMMAIRE

L'objet des techniques fondées sur la reconnaissance de la parole est double. En premier lieu, la maîtrise des objets directement à notre portée grâce à des dispositifs obéissant aux commandes orales. En second lieu, et c'est là l'objet principal de la présente recherche, la maîtrise des obstacles en dehors de notre portée, par le moyen de la communication sous toutes ses formes. Un coup d'oeil rapide sur les besoins ressentis par les personnes âgées et les handicapés en matière de services de transport améliorés montre que le besoin d'information est de loin le plus important.

Le présent rapport met en évidence un certain nombre de techniques fondées sur le traitement automatique de la parole, applicables en transport des personnes visées. Ce sont : les répondeurs automatiques à messages variables, les systèmes d'information personnalisée dans les installations terminales, les commandes orales de fauteuils roulants et les systèmes de guidage parlants. Afin de déterminer l'accueil que l'utilisateur pourrait réserver aux différents systèmes proposés, une recherche devra être faite sur cet aspect particulier.

L'étape suivante devra être celle de l'évaluation de chacune des applications envisagées : évaluation technico-économique et en service, obtenue soit par une recherche plus poussée, soit par une démonstration dans le cadre de programmes choisis. Le rapport fournit les critères en fonction desquels ces évaluations devraient être faites et qui visent à éclairer les voies que la recherche devrait suivre.

Le rapport conclut en formulant des recommandations sur le programme de R&D devant être lancé dans le domaine de la reconnaissance de la parole appliquée au transport des personnes âgées et des handicapés. Il donne plusieurs exemples d'applications, qui ne sont pas les seuls envisageables. Cependant, il rappelle l'existence d'un problème, celui de l'accueil que réservera l'utilisateur aux techniques proposées, toute innovation rencontrant une résistance tant générale que particulière qui devra être surmontée. Pour cela, il faudra soigner le mode de présentation et de mise en oeuvre et faire valoir certains facteurs valorisants : avantages escomptés, caractéristiques du produit comme la fiabilité, l'utilité, la robustesse, le naturel intonatif de la voix artificielle, etc.

Certains problèmes pourront être résolus par des perfectionnements techniques qui rendront les dispositifs proposés faciles d'emploi; d'autres nécessiteront une recherche plus poussée. Enfin, le rapport propose que les applications soient choisies en fonction des performances réalisées et attestées par des essais normalisés et compte tenu de l'effort requis pour la mise en oeuvre, d'autant plus que l'intervention de la parole risque de modifier profondément le comportement auquel s'étaient accoutumés les utilisateurs visés lorsqu'ils voulaient avoir accès au réseau de transport.

CONTENTS

1. INTRODUCTION 1
   - An Historical Perspective. 3
   - A Technological Overview. 4

2. SPEECH 5
   - A Uniquely-Human Ability 5
   - Listening With the Brain 7

3. TECHNOLOGY 9
   - What is Voice Recognition? 9
   - Voice Synthesis: Say That Again! 12
   - Voice Compression Technology 14

4. CONSTRAINTS 17
   - Why Doesn't It Understand Me? 18
   - Why Doesn't It Speak More Clearly? 22
   - Human Constaints on the Technology 24

5. CAPABILITIES 27
   - Current Status of Voice Recognition. 28
   - Voice Synthesis Capabilities. 31

6. APPLICATIONS 33
   - What You Can Do With the Technology 33
   - What the Disabled Have Done. 34
   - Some Opportunities in Transportation 35

7. DIRECTIONS 41
   - Assessing Voice Technology Applications 42
   - Recommended Course of Action. 44

REFERENCES 45

APPENDICES 47
   - Glossary 49
   - Vendors 53
   - Research Ctres 59
   - Systems. 61
   - Requirements 77

1. INTRODUCTION

The need to have control of our environment and control of our lives, in order to reduce dependence or even feelings of helplessness, is critical to human development. This may be especially true for persons with higher levels of disabilities. People with disabilities are often faced with both real and perceived loss of control of their environment and even their lives. These feelings of helplessness can be viewed as the antithesis of independence. A positive change in independence, self-reliance and life satisfaction requires that resources be used to focus on the challenges inherent in providing assistance that will enable all persons to function at their highest level of performance.

The Transportation Development Centre, in recognition of this concern for independence by individuals with disabilities, has commissioned a study into the application of voice technology to the transportation problems of the disabled. This study looks at voice technology as one of several related technologies encompassing computers, robotics, artificial intelligence and natural language processors that can be used to develop a positive change in the life styles of elderly and handicapped persons.

While technology is certainly not a panacea for answering all physical and psychological challenges, it is still a useful tool for addressing some of the shortfalls that individuals face when dealing with disabilities. For disabled people, voice-controlled devices could become an essential means to ameliorate lost functioning as well as an essential communication link with the outside world to augment lost mobility. Using technology in this way to serve human needs is not a novel idea. Using evolving technological innovations in this service role, however, is still an exciting new field of exploration.

INTRODUCTION

The speed and accuracy of the electronic computer is a well-known fact. Its cost is decreasing and its accessibility increasing with the advent of small personal computers. Effective use of these microcomputers, however, has always been limited by the ability to communicate with them. Input/output functions typically are the slowest aspects of a computer; therefore, any method that speeds up the ability to interact with a computer is worthwhile. The most efficient means of communicating among humans is speech, which is learned early in life. It is fast, convenient, and gets the maximum amount of information conveyed in the shortest amount of time. We are in fact speech-oriented. Why not have this same capability for speech an integral part of the machines we use?

The element of speech can add an entirely new dimension to the ways most people (especially the disabled) interact with machines, particularly the computer. As computers become more universally available and acceptable, the desire to communicate with them in more natural ways becomes increasingly desirable. Voice processing by computer provides one method for natural communication between people and machines. Although their potential is just beginning to show, a great deal is known about voice systems that recognize a small range of spoken commands and synthesize intelligible speech output. Moreover, the technology is not far off when natural spoken communication in continuous speech will be used for an increasingly complex range of applications.

The computer can be programmed to recognize voiced commands (voice recognition) and to provide voice output (voice synthesis). It can also be used to record, store and forward voice messages (voice compression). As input to a computer, voice technology can be used for entering data, for retrieving information, and for transmitting commands. Probably the most obvious use of voice recognition is in command and control, which in itself can take many forms, for example:

- Mobility Assists, particularly with voice-actuated wheelchairs, so that mobility can be achieved without hands or feet;
- Information Assists, obtaining standard information, particularly over the phone, without the need for clerks or operators;
- Life Style Assists, used by individuals to control various functions in their homes, affording them a fuller life;
- Productivity Assists, used by individuals in the workplace to allow them to perform particular tasks more effectively.

While voice recognition or voice input systems may appear to have a lot of applications, most advances have been made in voice synthesis or voice output technology. Most voice output products today are built from integrated chips or boards, which offer tremendous flexibility of design.

And since most voice chips on the market are quite good (in terms of vocabulary and intelligibility), the variety of 'speaking' products is growing daily. Designed with a speech recognizer, for example, speech synthesizers afford the elderly and disabled user a virtual 'human' interface to the powers of the computer.

As can be seen by the preceding brief introduction to voice technology, and as will be seen in greater detail in the following report, much work has been done over the last thirty years in developing computer-based voice communications.

AN HISTORICAL PERSPECTIVE

The idea of artificially produced speech and the ability of machines to recognize spoken words has fascinated researchers for many years. Mechanical speaking machines have been identified as far back as the eighteenth century. One of the earliest electrical speech synthesizers appeared in the late 1930s with a device called the Voder (Voice Operation Demonstrator). Developed by Homer Dudley at Bell Telephone Laboratories and exhibited at the New York World's Fair, the device modeled the human vocal tract via special filters. It led to the Vocoder, also by Dudley, which produced speech from coded electrical signals (1). Numerous artificial speech systems have been developed since then, probably none more useful to the visually-impaired than text-to-speech systems which create a human sounding voice from input text.

Speech recognition systems have a more recent history, with most of the early development work done in the 1950s. Prior to the work done in the '50s, a key development which paved the way for speech recognition was the sound spectrograph, developed in the 1930s and '40s, that produced a graphic picture of the energy characteristics of an utterance. The early recognition systems were oriented more toward individual word understanding while later developments in the 1960s and '70s began focusing on connected speech. One of the most significant voice recognition efforts undertaken was the Speech Understanding Research project of the Advanced Research Projects Agency in the United States. Lasting from 1971 to 1976, its goal was continuous speech recognition using a defined vocabulary, and was largely successful in meeting its objectives. Much of the work done in the Speech Understanding Research project and other projects formed the nucleus of the speech recognition industry today (2).

A natural offshoot of a voice synthesis system is a voice compression system. These can be either voice response systems or voice store-and-forward systems. Probably the most familiar of the voice response systems is a telephone intercept, in which a voice advises you if a phone number has been changed or disconnected. These systems date back

twenty years, and have been a mainstay of the telephone industry, particularly with improvements in computer technology. The area of voice compression technology receiving a lot of attention these days is voice store-and-forward systems. Using sophisticated computer switching and voice digitization and compression techniques, voice store-and-forward systems can send, receive, and retransmit voice messages to one or more users within the system (3).

A TECHNOLOGICAL OVERVIEW

As can be seen, many years of effort have been invested in the voice technology industry. Some areas, such as voice synthesis and voice response, are well established, with diverse product lines. Voice recognition, which tends to be more complex than other technologies, has also had its share of successes, although certain aspects, such as continuous speech recognition, still have quite a way to go before entering a real-world environment. Voice store-and-forward systems, in their current form, are relatively new, but are gradually finding their way into many business operations.

Computer-based speech is an increasingly important area in information processing, since it utilizes the form of communication most natural to people -- speech. This report examines three distinct areas: voice recognition, voice synthesis, and voice compression. The individual speech techniques will be reviewed, the current applications for speech examined, specific applications developed for improving the mobility of the elderly and disabled, and directions offered for planning and implementing a system. A highlight of the report is the detailed comparison charts featuring over a hundred speech products from about fifty vendors.

2. SPEECH

In order to understand how computers are able to speak and to recognize spoken words, it is useful to understand how humans speak and how they listen. By understanding the basic theory of speech production and recognition, it is easier to understand how computers can be programmed to both talk and listen. Very often, the acoustic analysis of speech involves undoing the sound so as to reveal the activities of the vocal organs that occurred when the words were produced; and synthesizing speech often involves using an analogue of at least some parts of the human vocal apparatus.

In this chapter, a brief overview will be provided of the uniquely-human ability of communicating through the spoken word.

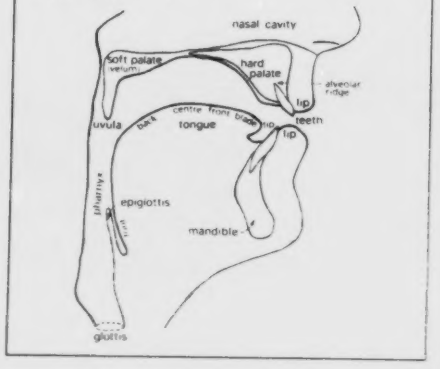
A UNIQUELY-HUMAN ABILITY

Although most of the higher animals generate a wide variety of sounds for communication purposes, it is the human race that has taken this capability to its most sophisticated level. Not content with just a series of grunts, clicks, and other primitive noises, humans found they could join these various sounds into more complex structures, which evolved into speech as we know it today. What we basically take for granted is actually the product of a complex grouping of elements, without which it would be impossible to produce speech.

The key element in speech, the vocal chords, vibrate to produce what are referred to as 'voiced' sounds, typically vowels. The other key element

in speech is the vocal tract, which includes the mouth, tongue, nasal cavity, and throat. All the other specialized sounds, such as consonants and consonant groups (i.e. ch, sh, gr, st, etc.) are produced in the vocal tract. In the formation of consonants, the airstream through the vocal tract is obstructed in some way. Add a healthy set of lungs and you have the primary ingredients for speech -- primary because the production of speech is only the first step in the communicative process. When we speak, we generally don't utter words at random, or with a relentless monotone. Our speech is generally well-ordered, and variations in speed, pitch, volume, and timbre give it additional meaning.

THE ORGANS OF THE VOCAL SYSTEM



The actual production of speech generates a complex series of rapidly changing pressure waveforms. Closer examination of these waveforms reveals the basic building-block elements of speech, known as **phonemes.** Roughly speaking, a phoneme is a group of similar, but not identical, sounds that differ from one another in accordance with the context in which each occurs. It is the basic unit for describing how speech conveys linguistic meaning. It is actually smaller than a syllable and even a half syllable. Each language has its own set of phonemes, and speech is comprised of phoneme groups. There are approximately 42 phonemes in the English language. Vowel and consonant sounds are, in fact, phonemes, and have different characteristics, depending on which components of the human speech apparatus are used to produce them.

Going a step beyond phonemes, we find that the position of a phoneme in a word has a significant effect on the ultimate sound generated. This implies the need for greater definition of specific speech sound, since on paper a word generally has a specific meaning. The same word spoken, however, can be perceived in different ways depending on how it's said. Speech scientists have identified **allophones** as 'acoustic variants' of phonemes and use them to more explicitly define how a particular word is formed. Allophones, like phonemes, are as variable as languages themselves, and it's not unusual for a basic phoneme to have one or more allophones derived from it -- such is the complexity of speech.

The basic components of words are **morphemes,** and generally take the form of prefixes, roots, or suffixes. Most words are comprised of one or more morphemes, which are themselves based on allophones and phonemes. Finally, **prosodics** include intonation, stress patterns, speed and all the other factors that give speech its 'human' feel. Without going into the

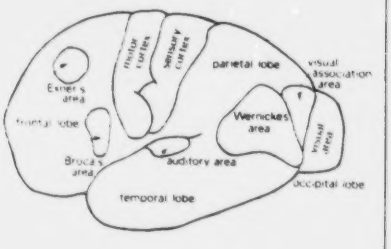
mathematical aspects of speech -- which are extensive -- it can be seen that speech is indeed a complicated process. And yet it is the most efficient means of communicating ever devised.

LISTENING WITH THE BRAIN

If speech is complicated, the human brain is even more so. It's the most complex piece of voice-recognition software available. Although our ears receive sound waves and convert them to electrical signals passed along the auditory nerves, it is the auditory center in the left temporal lobe of the cerebrum that gives us our sense of hearing. But distinguishing speech from other sounds is an even more specialized process.

Toward the posterior part of the left temporal lobe, between the auditory and visual regions lies the portion of the brain known as **Wernicke's area.** Since the last quarter of the 19th century, scientists have noted that patients with lesions in this part of the cerebral cortex cannot comprehend spoken language. They can still produce fluent speech sounds, although the meaning of their sentences is often quite jumbled. Wernicke's area appears to recognize words -- that is, it stores memories of word sounds and pairs them with ideas (4).

THE BRAIN'S SPEECH-RELATED AREAS



As in the rest of the brain, the mechanisms in Wernicke's area do not operate in isolation. 'Language' -- speaking, listening, reading, and writing -- is a multilayered ability involving motor skills, sensory reception, learning, and memory. Bundles of nerve fibers connect Wernicke's area to three other main language centers, all in the left hemisphere. Perception of written language is centered in the **visual-association area,** adjacent to the part of the cortex that gives us our sense of sight. **Borca's area,** located in the frontal lobe near the area controlling lip, tongue, and jaw muscles, coordinates our physical ability to talk -- it stores the memories of the muscle movements necessary to produce speech. **Exner's area,** near the part of the motor cortex associated with hand movement, regulates our ability to write language.

Normal perception of others' speech mediates the development of language production. Most models that attempt to explain speech cognition in the brain are variants of the **motor theory** of perception.

The motor theory of perception takes into account the fact that when we speak, we hear our own words, and when we hear others' speech, we unconsciously anticipate our own reproduction of those words. As we listen, we recall the ideas we were expressing when we pronounced the same words, and thus know what the other person means. Some studies have even detected minute movements in the larynxes of people listening to others' speech.

In contrast, computer voice recognition systems often 'recognize' speech by matching a voice pattern with a template stored in the computer. This straight template design of computerized speech-recognition devices does not seem to serve as a theory for human speech perception. We routinely understand different acoustic stimuli -- owing to wide variations in stress, accent, and intonation -- as the same phoneme. In this case, there can be no direct matching of sound waves to cognitive representations, or templates.

A compromise theory is the mechanical **analysis-by-synthesis** model. When we hear speech, we store those original auditory signals and generate hypotheses about their characteristics according to our own internal rules. We speculate: What would this input be like if our hypotheses were true? We then compare our descriptions to the stored input until we achieve a match. Various techniques such as linear predictive coding, dynamic programming or even straight statistical analysis are being applied to develop better analysis-by-synthesis models of human speech perception.

Cognitive theorists acknowledge the vagueness of the current methods of analyzing human speech perception. The models barely scratch the surface of the complex of abilities called human cognition. Within their constructions, speech scientists can not yet account for even the comparatively minute amount of biochemical and physiological data that has been gathered. Computers as tools may someday help us sort out the mystery of the human mind -- but not even the most sophisticated machine can duplicate it.

3. TECHNOLOGY

The preceding chapter was intended to provide a very basic appreciation of the complexities of the human speech process: the ways in which the vocal tract produces sounds and the ways in which the human mind perceives the sounds. Let's now have a look at how electronic devices are attempting to duplicate the process.

WHAT IS VOICE RECOGNITION?

Looking at the human speech process makes you stop and think how good we are at recognizing and understanding speech. It's an ability that we usually take for granted, and we do it quite well. For a computer, however, speech recognition is a much more complicated process. It can be defined as the ability to take speech input, and respond with the desired output. This rather simplistic definition belies the enormous complexities inherent in speech recognition. The problem is further complicated by the diversity of human vocal capabilities. After all, it's been shown that no two voices sound exactly alike. Add to this the elements of context and semantics, which imply additional meaning to what is said, and you have a real challenge.

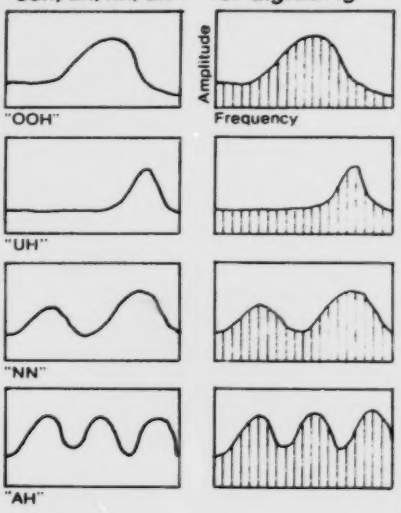
The majority of voice recognition systems attempt to recognize isolated words, and, by and large, do that quite successfully. Only a few systems can recognize groupings of words, or continuous speech, and are generally limited in their capabilities to only a few words per statement. Again, most systems are speaker-dependent, being programmed to respond to specific voice patterns embedded in memory. Much more difficult is

THE VOICE RECOGNITION PROCESS

Training a Voice Recognition System to hear "ONE"

❶ *Word "ONE" consists of four separate sounds "ooh, uh, nn, ah".*

❷ *Filters sample each sound at small intervals for digitizing.*



❸ *The sampled amplitudes are stored in a matrix.*

Sound	Frequency in Hundredths of a Second									
	1	2	3	4	5	6	7	8	9	10
"OOH"	15	16	17	18	20	23	30	35	38	37
"UH"	12	12	12	13	13	13	15	15	17	20
"NN"	18	25	30	25	15	12	11	12	17	26
"AH"	11	27	32	24	18	26	39	31	22	16

❹ *Training establishes a three-part entry in a vocabulary file, "Numbers.Voc".*

NUMBERS.VOC		
Command	Keystroke	Matrix
ONE	1	15, 16, 17, 18, 20...
TWO	2	26, 41, 28, 29, 42...
NINE	9	14, 17, 18, 19, 20...

speaker-independent recognition, which means the system must transcend the speech boundaries of a few known people and be able to understand many unknown ones. Speaker-independent systems require much more processing power and generally have a much smaller vocabulary than speaker-dependent systems. Both dependent and independent systems must also be able to screen out extraneous noises that could complicate the recognition process. Totally speaker-independent and fully-comprehending systems are still a long way off.

A computer 'recognizes' speech by a process of pattern-matching, comparing its analysis of what you are saying with the patterns of words it already has on file. A single word is easy to recognize because it is 'discrete speech' -- it has a clear beginning and end, and therefore a clear pattern. A string of several words is more difficult to identify; before the computer can compare words with its database, it must first decide when the sound level has dropped low enough to justify designating a break between words. The computer analyzes the entire multiword pattern, deciding on the basis of preprogrammed statistical reasoning when word breaks are most likely to occur. It will then take these designated words and try to match them with the patterns on file.

Speech recognition systems are based on several distinct components, as shown in the accompanying figure. A word is spoken into a microphone where it is converted to an electrical signal. The signal is then processed through an analog-to-digital converter and stored in the system. The signal is sampled from 6,000 to 20,000 times per second and each sample contains

between 6 and 16 bits of information. This data is then analyzed for specific acoustic properties that will identify the individual speech components. Several techniques are used in this process, such as zero-crossing, Fast Fourier transform, and linear predictive coding (LPC). Once the signal has been analyzed, the system will need less computational power when it comes time to match the signal with a reference template. Once a discrete speech signal is identified, it is then compared against a reference template library for a possible match. The systems must have templates programmed into memory. This is how the system is 'trained'. Since the likelihood of a perfect match is slim, many templates are programmed, giving the system a better chance to find a match. If a match occurs, the system will perform the desired action, such as printing what was spoken (5). In isolated-word systems, this process generally produces a high recognition rate, typically above 95 percent. In continuous speech recognition systems, the signal must be matched not only against a template, but also against a language model. This model describes the rules for the permitted vocabulary and the creation of word strings (syntax). As might be expected, language models for speech systems are not as complex as those we use every day, but that is generally due to the system's limited processing power.

The matching of human voice to computerized patterns is usually pretty close, but sometimes not close enough. A speaker's gender, degree of fatigue, rate of speech, loudness, or a head cold can change words' pitch or stress. Because of these factors, the lowest error rates occur in discrete-word, speaker-dependent applications.

After training, the spoken "ONE" is compared to entries in the "Numbers.Voc" file

❶ *The converter digitizes the spoken word "ONE"*

SPOKEN ONE	14.16.17.17.20

❷ *The program looks for similar entries in the file.*

NUMBERS.VOC		
Command	Keystroke	Matrix
ONE	1	15.16.17.18.20...
TWO	2	26.41.28.29.42...
NINE	9	14.17.18.19.20

❸ *The converter compares the spoken word "ONE" to similar file entries.*

SPOKEN ONE	14.16.17.17.20
ENTRY ONE	15.16.17.18.20
ENTRY NINE	14.17.18.19.20

❹ *The program chooses the closest match, the entry for number "ONE".*

❺ *The program sends the spoken entry "1" to the computer.*

❻ *The "1" is displayed on-screen.*

THE VOICE SYNTHESIS PROCESS

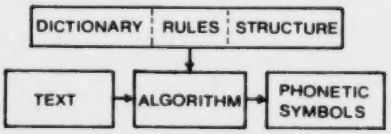
Getting a Voice Synthesis System to say "Two plus seven is less than ten."

❶ *Type in "Two plus seven is less than ten".*

❷ *The sequence is recognized with all its ambiguities.*

❸ *The microprocessor (the brains) begins the process of speech synthesis.*

❹ *The words are shaped and molded.*



(Text-to-speech algorithm applies a set of rules to words so that a sequence of phonemes are generated.)

❺ *A phonetic sequence is then generated.*

tuplʌsɛnɪzɪzlɛsðɔntɛn

(Durations and timing along with basic sound frequency are added.)

VOICE SYNTHESIS: SAY THAT AGAIN

Artificial speech is a seasoned technology with many examples already in existence. While most speech produced by today's technology still sounds distinctly non-human, the day is not far off when you won't be able to distinguish it from the real thing.

Speech synthesizers use several different methods for constructing speech, as illustrated in the accompanying figure. The most basic is known as digitized speech. It involves recording human-spoken words whose wave forms are digitally encoded and stored. To produce a word or phrase, the synthesis program retrieves the desired words from memory, assembles them in the desired sequence, and sends them to a digital-to-analog decoder for generating the voice signal. This method, although using incredible amounts of computer memory, is fairly simple and inexpensive. But the reconstructed speech exhibits little conformity to contextual or semantic rules. Two other methods of speech synthesis that reduce the memory requirements while still producing good quality speech are pulse code modulation and delta modulation. They use a technique similar to digitized speech, except that the process compresses the speech signal (6).

Another technique is linear predictive coding (LPC). Instead of duplicating the speech signal as in the previous waveform encoding techniques, the human vocal tract is mathematically modeled. Sounds are generated and language rules are programmed into the generated sounds to reproduce speech. It capitalizes on how the speech waveform is produced rather than merely duplicating it. The name, linear

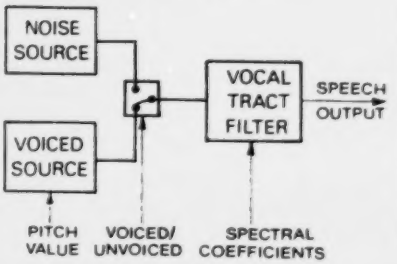
predictive coding, is derived from the technique used: it predicts the next coded speech sample using a linear combination of the preceding speech samples (7).

Still another technique, similar to waveform encoding, is called formant synthesis. Words are spoken into the system, where they are analyzed and stored as time variations of vocal tract resonances, called formants. To construct speech, a special digital filter, under computer control, retrieves both voiced and non-voiced speech components and assembles them in the correct sequence. This method produces much smoother flowing speech, and more attention can be paid to speech characteristics such as pitch, duration, and intensity. Both this method and waveform encoding have been used for producing the familiar telephone intercept message.

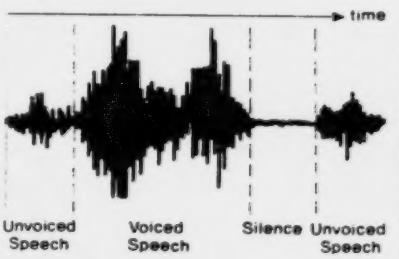
A relatively new synthesis technique uses digitally-encoded phonemes and allophones to construct speech. The method is very similar to formant synthesis, and has the capacity for an unlimited vocabulary. The input is broken down into phonemes, however, the resulting speech is synthesized largely from allophones, which can convey more contextual information than phonemes only. Allophones are produced by 12-bit control words acting on a preprogrammed phonemic synthesizer. Specific word characteristics (pitch, tone, volume, rate) are determined by the control words as they are decoded into analog control signals. The system hardware stores, the rules, or parameters, for synthesizing word forms. These parameters are then translated into speech sounds through an array of sound generators and filters. A

After the computer analyzes the rules, the speech chip takes over and produces sound.

❻ ***The speech chip now models or mimics the vocal tract.***



❼ ***The speech output is then transformed to analogue signals for speech.***



❽ ***The signals are fed into an audio amplifier and then into a speaker.***

❾ ***The speaker produces the words.***

tu plas ɛvən ɪz lɛs ðən tɛn

(phonetic transcription)

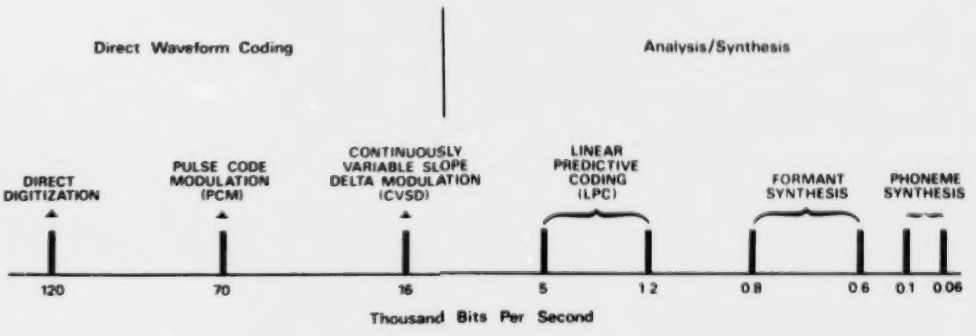
(21 sounds are generated but actually only 14 phonemes are used, with some used more than once.)

variable-pitch generator produces voiced sounds, such as vowels, and a white-noise generator creates unvoiced sounds.

Text-to-speech synthesis is a promising technology, similar to formant synthesis, in which typed text is the input medium. The system converts alphanumeric characters to speech, and also features an unlimited vocabulary. The text is compared with the system's dictionary, and if a match is found, the phonetic information needed to construct the word is sent to a formant synthesizer, from which speech output occurs. The system dictionary can also include specific acoustic characteristics (pitch, volume, duration, etc.) that will produce a more natural sound.

The figure below illustrates the basic difference between the various techniques that are used to synthesize speech (8). As can be seen, the amount of memory and computer processing power that is required varies considerable between the various techniques.

TECHNIQUES OF SPEECH SYNTHESIS



Summary of the different techniques of speech synthesis and their related computer memory storage requirements in terms of bits of memory required to store one second of speech. Notice the drastic reduction in memory requirements from the PCM technique to the phoneme technique.

VOICE COMPRESSION TECHNOLOGY

A logical application of voice synthesis technology is in the area of computer-based voice compression systems. Voice compression involves the conversion of analog human speech to compressed digital form for storage, and the subsequent decompression and conversion of the digital form to analog for playback. Since voice compression is an actual digital recording of human speech, the intonation, mood, and presence of the speaker all are accurately maintained.

As with recognition and synthesis, significant algorithm and technology advances have been made with voice compression. These advances have led to a much wider ranger over which human speech can be compressed without impairing the quality of the information being stored. This in turn has led to improved voice quality and reduced storage requirements.

Voice store-and-forward and voice-response systems utilize voice-compression technology. In a typical system, a user performs some sort of data entry function using a keyboard or touch-tone telephone. At the completion of the data entry action, the system replies with a specific response. A vocabulary of words and phrases is programmed into the system, using various technologies mentioned earlier. Upon receipt of a specific code, the synthesis program assembles the correct response from the vocabulary, where it is then converted to speech. The most familiar uses of voice response technology are in the banking industry, with a growing emphasis in electronic funds transfer and personal banking.

In most voice applications, users are prompted by static or fixed directives from the system. If these directives are recorded speech, the application generally is called voice response. If the system enables a user's recorded speech to be constantly changing or dynamic, the application generally is called voice store-and-forward. Voice messaging/mail systems provide both voice response (recorded prompts guide the user through the messaging process) and voice store-and-forward (the user can create, store, send, and receive messages).

Voice response systems are usually connected to telephone systems, since the telephone (particularly with Touch-Tone) is a natural data entry device. It's rather easy to connect a phone system into a company's computer; from that point voice response can be added and you're in business. The voice system will have a number of communications ports that connect to similar ports on the phone system.

4. CONSTRAINTS

The design and construction of systems for speech communication between man and machines is unquestionably a fascinating subject. There has been a great interest in synthesizing speech from some type of symbolic (usually textural) input, the 'recognition' of speech by means of transforming it to some symbolic form (sometimes followed by consequent actions), and the efficient coding of speech for compact storage and playback. Research on these systems has provided a great stimulus to many areas of science and engineering, including linguistics, phonetics, information theory, system modeling, signal processing, computer science, and technology.

The application of technology to speech communication provides excellent examples for the study of complex systems, since it raises fundamental issues in the choice of descriptive units, subjective techniques for the measurement of quality, naturalness and preference for speech, formalisms for representing human knowledge, representational techniques, and the exploitation of contemporary technology to produce real-time performance with acceptable cost. In addition, studies of talking and hearing have demanded ingenious instrumentation techniques, and a rarely-precedented level of system modeling, generation, and parameterization that can only be characterized by a very high degree of difficulty.

Continuing study has revealed a few first-order successes in voice technology. But there are unquestionably huge areas of constraints that must be resolved before there can be any degree of optimism for matching the performance exhibited by nearly all humans (9). In this chapter, we will look at some of the issues and constraints being addressed in current voice technology research.

WHY DOESN'T IT UNDERSTAND ME

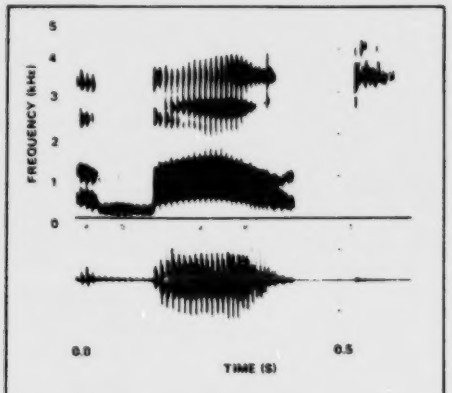
Speech Connectivity

Computer identification of word boundaries in natural speech is a difficult problem and a significant source of recognition error. Therefore, some systems employ isolated speech, which requires the user to self-consciously introduce short pauses between each spoken word. This provides increased accuracy at the cost of some inconvenience to the user. In some systems, this interword pause can be made so short that the data entry speed approaches that of natural speech.

The term connected speech refers to carefully spoken phrases without interword pauses. The number of words that can be spoken before a pause is limited. Continuous speech refers to unconstrained spoken sentences of unlimited length. This term also implies the use of a flexible syntax that approximates the freedom of natural language.

Most currently available recognizers depend on a small period of silence (typically 200 ms) between words to determine the end points of the word. These recognizers are clearly **discrete-speech** recognizers. Some recognizers are capable of recognition without explicit knowledge of the word end points. This technology tends to be expensive because it requires much more intensive data comparison than if the word end points were known. Even these **connected-speech** recognizers, however, do not perform as well with connected speech as they do with discrete speech because the acoustic variations of words spoken in connected speech is greater. This is attributable to the coarticulation of neighbouring sounds; the positions of the tongue, jaw, and lips in one speech sound are affected by their previous and future positions.

Of all the problems that burden the use of current speech recognizers, probably the greatest is this difficulty in reliably determining word endpoints. The spectrogram in the accompanying figure illustrates some of the problems in end-point detection. End-point misalignment is responsible for almost all word-recognition errors in the typical operation of discrete word recognizers that base reference comparisons on an initial determination of end points. This end-point detection problem does not exist (or rather, is solved in a more comprehensive way) for connected-speech recognizers. Therefore, connected-speech recognizers



End-point detection is difficult with low amplitude signals. This spectrogram is of the word "about" and shows the potential end-point detection problem.

have a potential for better performance even in discrete-speech applications (10).

Speech sounds have a wide range of amplitude and time profiles, causing difficulty in discriminating them from noise. There is also considerable variation in the way a speaker articulates a word; he will sometimes articulate a final 't' and sometimes not. In addition, the noise includes aspiration after speaking a word and quick inhalation or lip pops immediately before speaking. This does not happen on every word, but often enough to cause problems.

The most disastrous time for mistakes in end-point detection is during speaker enrollment. This can and often does create a rather common condition that is known as bad enrollment. When this occurs, the speaker must be reenrolled for the offending word. The problem is not so much that reenrollment is necessary, but rather that the need is not known until sufficient negative results are experienced to mandate the reenrollment.

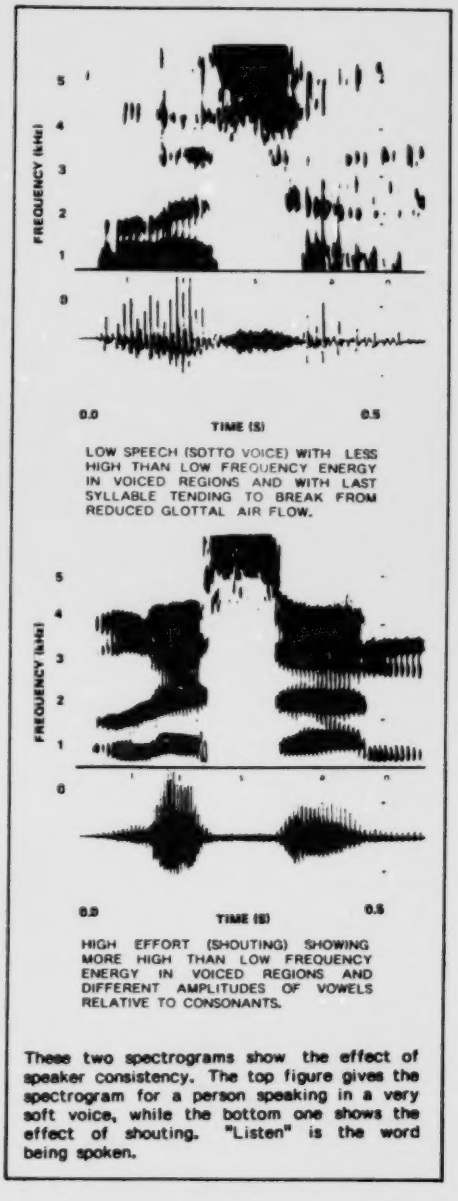
Speaker Consistency

Running a close second to end-point detection problems are problems in the consistency with which a speaker pronounces words. Indeed, inconsistency contributes to the end-point detection problem as well. Many physical and psychological factors contribute to variations in a speaker's pronunciation. Although little is understood about how such factors affect speech, we observe that speech data in a single session is much more self-consistent than speech data produced over a long period of time. The source of this variation may not be totally attributable to the speaker; other sources include variations in the environment (noise and reverberation) and in the microphone placement.

A speaker may improve the acoustic consistency by controlling speech effort level. The dramatic changes that take place with changes in speech effort are illustrated in the figure on the following page. Usually, the tendency of a new user is to speak much too softly. (Due to the fact that the speaker may be intimidated by the system and due to the location of the microphone, which is usually positioned no further than an inch or two from the lips, implicitly commanding a very low speech level). Low speaking levels promote inconsistency because of irregular voice excitation. A low speech level also reduces signal-to-noise ratio. Shouting would cure these problems and would be acceptable, except that consistency would be rather short-lived before hoarseness of the user and general disruption of the social environment occurred. The preferred speech level is that of an authoritative, confident, across-the-desk conversation. If this speech level is maintained, the new user has made an auspicious beginning.

The performance of machine word recognizers is also affected by limitations in the reliability of human speech performance. Similar problems are occasionally experienced in conversation with other humans, and these problems are solved through intelligent interactive feedback. These problems are solved in a like manner for machine recognition. In this case, the reliability of the human is of greater concern because of the larger set of rules that he must follow and, therefore, the greater the opportunity for making mistakes. Probably the greatest reliability problem for humans is the word separation required by discrete-speech recognizers. In the first place, we have observed that speaking discretely is not necessarily a skill that is intuitive or trivially acquired. One must learn to speak crisply and to leave the required gap (plus a safety factor) between words.

In the second place, the implicit demand for high recognition throughput requires that his safety factor be no greater than necessary. This can only be learned through experience. With this perspective, it is clear that discrete-speech errors are an important factor in the performance of discrete-speech recognizers. Connected-word recognizers, by contrast, have the advantage of a gradual degradation of performance with decreasing size of the gap.



LOW SPEECH (SOTTO VOICE) WITH LESS HIGH THAN LOW FREQUENCY ENERGY IN VOICED REGIONS AND WITH LAST SYLLABLE TENDING TO BREAK FROM REDUCED GLOTTAL AIR FLOW.

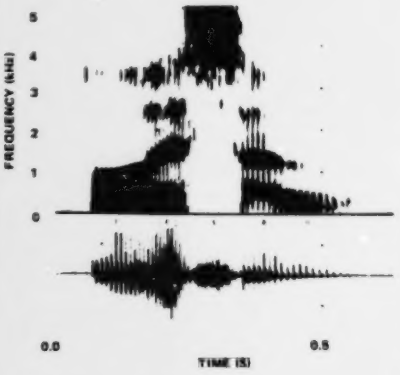
HIGH EFFORT (SHOUTING) SHOWING MORE HIGH THAN LOW FREQUENCY ENERGY IN VOICED REGIONS AND DIFFERENT AMPLITUDES OF VOWELS RELATIVE TO CONSONANTS.

These two spectrograms show the effect of speaker consistency. The top figure gives the spectrogram for a person speaking in a very soft voice, while the bottom one shows the effect of shouting. "Listen" is the word being spoken.

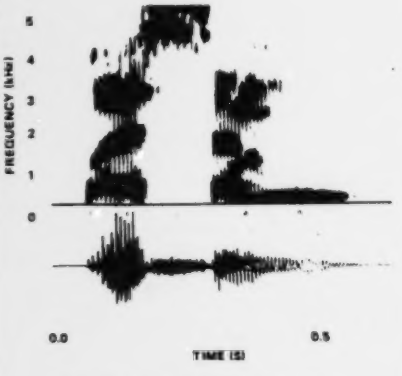
Speaker Dependence

This is a measure of the amount of speech information each user of the speech recognition device must provide to obtain favourable results. **Speaker-dependent** systems generally require the user to train the recognizer by repeating each word in the vocabulary several times, usually between one and ten. (the recognition accuracy generally increases with each training repetition, up to about five.) Such systems then form word templates

from these training utterances. **Speaker-independent** systems attempt to accommodate the speech of a large class of users, without a training process. Both approaches may also include an adaptive mechanism that updates the template data base as recognition proceeds in order to achieve better performance and dynamic characteristics. Some speaker-independent systems may require several sample sentences from the user to initially align certain recognition parameters.



A SPECTROGRAM OF THE WORD "LISTEN". THE BEGINNING AND END OF THE WORD EXHIBIT LOW-FREQUENCY ENERGY, WHILE THE "S" EXHIBITS HIGH-FREQUENCY ENERGY.



A SPECTROGRAM FOR THE SAME WORD PLOTTED FROM A NORMAL LEVEL SPEECH SIGNAL FOR ANOTHER SPEAKER SHOWING A HIGHER AMPLITUDE AND SHORTER TIME COURSE.

Performance is degraded when speaker-independent recognition is being attempted. These two spectrograms are for the word "listen" but for different speakers. Note the different formant amplitudes and time course.

The second factor to consider is the statistical distribution of performance as a function of speaker over the user population. Performance varies radically from speaker to speaker, and performance for speaker-independent recognizers really needs to be characterized in terms of population performance statistics. Not all users will be equally successful in using currently available speech recognizers. This problem occurs mainly in the context of speaker-independent recognition, but the same problem occurs even with speaker-dependent recognizers. Differences in performance for various users may be partially attributable to the experience of the user, but beyond this a very strong individual variation in recognition performance still remains. This variation has been observed many times. Although a continuous spectrum of performance occurs across all users, it appears that recognition systems generally work well for the bulk of the population and that most problems are created by a small population segment.

In general, the performance of speaker-dependent systems is superior to that of speaker-independent systems. In addition, speaker-dependent recognizers can usually accommodate a much wider range of language and dialect than their speaker-independent counterparts. The obvious disadvantage of speaker-dependent systems is the inconvenience of repeating each word in the vocabulary several times in order to train the device.

Vocabulary Size

The size of the application vocabulary also affects recognizer performance. As the vocabulary expands, the probability of acoustic error increases and the recognition accuracy decreases. In addition, the recognition processing time generally rises with increasing vocabulary size because the recognizer must consider more possibilities. These effects can be partially controlled by dividing the application vocabulary into smaller sub-vocabularies, which can be in effect at different points in the application task. Obviously, the vocabulary size for speaker-dependent recognizers is limited by the need to repeat each word. However, a surprising number of practical applications can be automated with small vocabularies, often containing only 50 to 250 words.

Of course, capability is not strictly a function of vocabulary size. The real issue is how similar the words are in the vocabulary. Performance tends to be better for vocabularies consisting of long multi-syllabic words. Most important is the particular composition of the vocabulary in terms of the acoustic similarity of the various words to one another. For example, even a two-word vocabulary might be impossibly difficult (such as 'seen' and 'seem', or 'one' and 'won'). These similar sounding words are a problem that cannot by solved by acoustic techniques alone; al limited vocabulary must be selected to eliminate homonyms, or such clues as syntax must be used by the computer. The words do not have to be homonyms to cause the problem; many systems have difficulty differentiating between the numbers 'five' and 'nine' (11).

Another human limitation relates to the size of the recognizer vocabulary. As the vocabulary size is increased, there may be a problem in the human recollection of the vocabulary. This is particularly true if portions of the vocabulary are seldom used -- a frequent situation for large vocabularies. This human limitation may be overcome by experience or by visual displays, but it is a problem that needs to be considered by potential users in judging the suitability of their intended applications.

WHY DOESN'T IT SPEAK MORE CLEARLY

Units for Speech Processing

As was mentioned in the previous chapter, speech synthesis depends on the processing of basic units of natural speech to serve as models for the synthesis system. There are clearly trade-offs here between the size of the basic units, the capability of the system to deliver a large range of messages, and the resulting quality of the synthetic speech. Thus at the small end, there are relatively few phonemes, but as the units become larger through diphone, demisyllable, syllable, word or phrase, the number

of items of the unit grows without limit (and the quality of the speech output improves). On the other hand, the smaller the unit, the more abstract it is likely to be (12).

Generally then, the larger the basic unit, the more natural the synthetic speech will sound, owing to the fact that some effects of word stress and coarticulation are retained in the larger units of speech. If the basic units are as large as words, however, the synthesis system must be capable of storing thousands of encoded speech samples in memory in order to deliver the range of messages that the 'phoneme' system is capable of, because the phoneme system can be used to build any message from a repertoire of approximately 50 phonetic segments. By stringing phonemes together, any word in English can be reproduced.

As a result of this factor, there is a general tendency towards increased abstraction (and, hence smaller units) as additional knowledge about speech becomes codified. Composing speech using smaller basic units places a tension between internal and external structure and cohesion. Thus the phoneme has relatively little internal structure, but considerable external structure is needed to specify the transitions between phonemes at their boundaries. Diphones and syllables, on the other hand, have a richer internal structure, and require less specification at their boundaries. The search for the natural 'joints' of language will certainly continue in an effort to determine the optimal units.

Coarticulation Effects

Part of the reason for searching for the natural joints of language is the effects of coarticulation. The effects of coarticulation in natural speech is such that one phonetic segment may influence another that is up to five or six segments away. This influence is not restricted to within a syllable or word, so systems using even these larger units of speech are not necessarily free of the effect.

The task of producing natural sounding speech by linking the basic units of speech together is therefore complicated by the influence, or lack of influence, of one unit upon its neighbour. This is the primary source of the unnatural qualities that often come from synthetic speech. Procedures for smoothing the signal across unit boundaries are often employed, involving interpolation of parameter values to modify the acoustic quality of one segment to be similar to its neighbour. However, these procedures are generally too simplistic to handle the complex interactions of phonetic segments. This is because the procedures do not take into account the degree of dominance of certain articulatory gestures over others, or to adequately anticipate the temporal aspects that govern rules of coarticulation. Further research is therefore required to obtain solutions to these problems.

Stress and Intonation Rules

For the generation of natural-sounding synthetic speech from small units of speech, rules that provide the correct emphasis to sentences are required. Often the meaning of a sentence depends solely on the placement of stress and intonation. Stress is realized acoustically as variations in the relative duration, amplitude and fundamental frequency of syllables, while the primary cue for intonation is found in changes in the fundamental frequency. Different strategies are employed by different languages, and linguistic theory indicates that there are word, clause and sentence levels of stress and intonation.

Some text-to-speech systems have been fairly successful in applying stress and intonation rules, however, the range of these rules is limited, and it may depend upon the rapidly developing technology of expert systems to provide the means of producing synthetic speech that is capable of reflecting subtle differences in meaning, intent and mood.

HUMAN FACTOR CONSTRAINTS ON THE TECHNOLOGY

Human Stress

The first rule in designing any human-machine interaction is to make the machine fit the human, not the other way around. Unfortunately, accurate voice communications between a human and a machine requires additional effort to speak (and listen) clearly -- thus placing stress on the user. To articulate clearly and the ability to do so is degraded by fatigue and it is further degraded when the speaking rate is increased.

Stress in voice systems is of two kinds. One is based on unfamiliarity with the system and may go away with practice, but the other does not. If the system is being used extensively, fatigue becomes an important constraint. After a while, the operator may not speak crisply or loudly enough. Often groups of words may be slurred, which creates especially severe problems for isolated-speech recognition hardware.

Noise

With discrete-speech voice recognition systems, background noise can mask the pause between words, so that a noise-cancelling microphone is essential where noise is high. But with such microphones, breath noise is introduced which also interferes with the boundary between words. Unfortunately, with a noise-cancelling microphone, a shift of a few degrees between speaker signal source and location also effectively degrades the cancelling feature.

An office or public environment presents particular problems. In an office, for example, typewriters, printers and conversation can interfere with intelligibility. In a public place, an operator often will not speak up loudly enough in order not to offend passers-by.

Accuracy

Voice recognition and voice synthesis will rarely achieve an accuracy of 100 percent in a typical application. Accuracy constraints must be assessed in terms of the consequences of an error. Consideration must be given to the how critical accuracy is to an application and how errors are detected and corrected.

Unless the accuracy is very good, the operator may lose confidence in the system, stress will increase, and performance will do down further.

Training and Retraining

There are two types of training with voice recognition systems: The speaker trains the system to understand him and the system trains the speaker in what vocabulary to use. There are also two phases of training, the initial period and subsequent retraining.

The operator trains the system in his method of speaking each word. This requires repeating the vocabulary list several times. The speaker learns that the best performance is achieved by speaking crisply, clearly, quickly (but not exceeding the maximum speech rate) with rhythm, and at normal loudness. The most common error in training is to overpronounce to try to help the machine. Later on, when using the system, the operator speaks more normally, at which time the recognition system starts to make errors. After using the system for a while, slurring or errors in pronunciation will appear, requiring the system to be retrained. In some cases, the whole vocabulary may have to be relearned by the system.

Technological Push

It is the nature of technology today that advances occur even if there is limited need for the technology. As voice recognition and synthesis devices are developed, they are incorporated into numerous kinds of systems. The fact that many of these applications are of little value does not really matter. We are going through a period of 'technology push', where the technology is available and introduced even without consideration of costs and benefits. Indeed, the problem is sometimes to determine uses for the technology already developed (as well as to keep up with the technological changes).

Increased Complexity

The quest to make tasks easier and more efficient and increase productivity has often resulted in solutions that have precisely the opposite effect. Certainly, we have capabilities now that far exceed those available before technological improvements occurred. Yet many routine things are more difficult to perform on an automated system. Yet we come to rely on the new technology and use it for all tasks, even where inappropriate.

Health and Safety

The introduction of new technology always raises a number of concerns about health and safety. It is a certainty that some of the changes that may be made in the transportation field will become health and, particularly, safety issues.

5. CAPABILITIES

There are many unresolved issues in voice technology, but research and development in the area is growing rapidly. While the major companies in the industry are small- to medium-sized. the industry leaders, such as Hewlett-Packard, IBM, National Semiconductor, Texas Instruments, and Wang Laboratories all have R&D programs well under way. Extensive work is also being done in universities and government agencies. Most of the work centers on producing more powerful, less costly speech components with larger vocabularies, better recognition capabilities, and more natural-sounding speech.

The speech recognition industry is searching for the biggest prize -- a speaker-independent continuous-speech system. Most experts agree this system is still years ahead, unless a significant technological breakthrough occurs. In the meantime the speech recognition industry is intent on improving their systems' understanding capabilities, ease of 'training' and vocabularies, while making them more cost-effective and easier to use. Improvements in semiconductor technology are expected to contribute to these goals.

Until recently, voice synthesis systems could be classified into two broad categories depending on the nature of the synthesis process. Speech-coding systems used a fixed set of parameters to reproduce a relatively limited vocabulary of utterances. These systems produced intelligible and acceptable speech at the cost of flexibility in terms of the range of utterances that could be produced. In contrast, synthetic speech produced by rule provided less intelligible and less natural sounding speech, but these systems had the capability of automatically converting unrestricted text into speech. Over the last few years, significant improvements in

text-to-speech systems have begun to eliminate the advantages of simple coded-speech voice response systems over text-to-speech systems. Extensive research on improving the letter-to-sound rules and phonetic implementation rules used by these systems suggest that, in the near future, unrestricted text-to-speech voice response devices may produce highly intelligible and very natural sounding synthetic speech.

CURRENT STATUS OF VOICE RECOGNITION

One of the most ambitious voice recognition systems comes from a relatively new company, Kurzweil Applied Intelligence (Waltham, Massachusetts). The company is the latest high tech entrepreneurial effort of Raymond Kurzweil, who earlier founded one company that makes text-reading machines and another that makes electronic musical keyboards capable of emulating an orchestra.

In the fall of 1985, Kurzweil began shipping the KVS-3000, which the company claims has a speaker-dependent vocabulary of 1000 words and the ability to interface transparently with almost any personal computer or mainframe. Kurzweil is touting the KVS as a forerunner to a more flexible system called the Kurzweil Voice Writer, a 15,000-word voice-activated typewriter that is also intended to be compatible with the IBM PC and other popular business computers and software. The Kurzweil Voice Writer will be an adaptive speaker-dependent system, which means that the system will continue to collect data about the user's voice and speech habits even after the initial training session has been completed. It will thus become increasingly customized to the speaker, and accuracy should continue to increase over the time a person uses the system, eventually exceeding 95 percent (13).

In addition to matching sound patterns, the Kurzweil Voice Writer (KVW) will try to anticipate the next word in a sentence; linguistic analysis of a sentence will tell what part of speech (noun, verb, adjective, etc.) to expect. Although Kurzweil hopes to bring out the KVW in 1986, significant improvements are still needed. Speed, for example, must be increased from the present pace of two minutes per word to something approaching real time. Implementation of a new architecture and custom VLSI (Very Large Scale Integrated) circuits are proposed to cut recognition time to an average of 250 milliseconds.

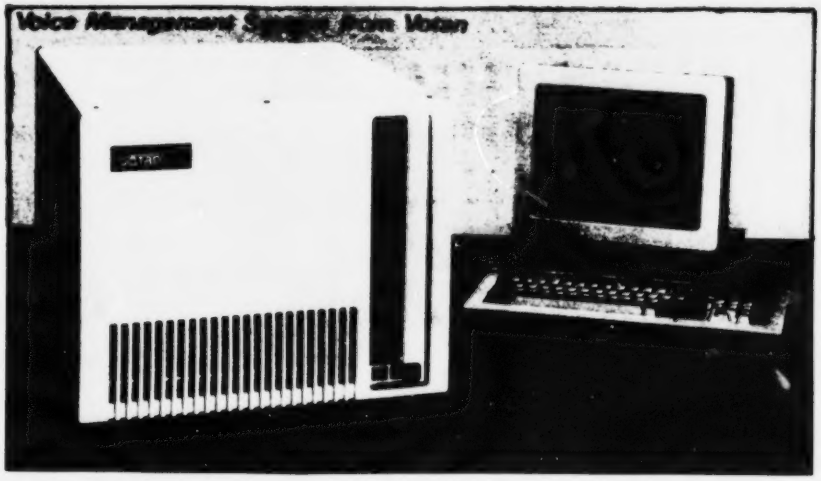
Another large-vocabulary (5000-word) voice recognizer is promised for later in 1986 by Speech Systems (Tarzana, California). Unlike Kurzweil, Speech Systems did not first produce a limited-capability machine in order to test the market. The company's Voice Line will be its first major product and it will combine statistics and artificial intelligence to yield a high degree of accuracy in voice recognition. So far, the company has

demonstrated recognition of hundreds of words, but only at about one-seventh the speed needed for real-time dictation.

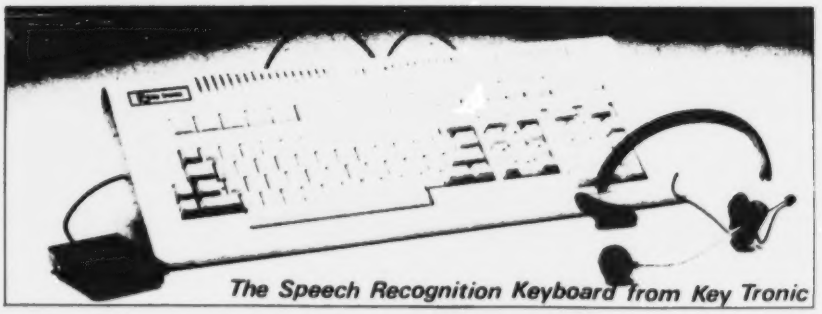
Despite its considerable research and development in the field, IBM has yet to announce a commercial voice recognition system. However, the company has demonstrated an experimental 5000-word system at its Thomas J. Watson research center (Yorktown Heights, New York). The experimental system achieves an accuracy of 95 percent for its vocabulary of common office terms. Like Kurzweil's planned Voice Writer, the IBM machine tries to anticipate the next word. But IBM takes a statistical rather than linguistic approach. Once the system identifies a few words, it computes the probability of a given word coming next, using as a reference guide some 25 million words of office correspondence that it has previously analyzed. The system is limited to discrete words and must be retrained for each new speaker. IBM is trying to shrink the present mainframe-based system into a box that could be connected to a Personal Computer, but no such product is imminent.

IBM did introduce a more limited voice recognition product in the fall of 1985, though. Using technology licensed from Dragon Systems (Newton, Massachusetts), the unit has a 64-word vocabulary and can be upgraded in 64-word increments by sacrificing response time. The system can be used to enter commands and data for any software that runs on the MS-DOS operating system (the standard operating system for the IBM PC).

Probably the largest single supplier of voice recognition hardware has been Votan (Fremont, California), which makes a voice input board for the IBM PC as well as several stand-alone units for voice mail and telemarketing, and for input to IBM mainframe computers. Votan says its products use proprietary software to recognize vocabularies of several hundred words in continuous speech, without the pauses that most systems require to determine when a new word begins. Votan recently agreed to supply voice recognition boards to Sperry (Blue Bell, Pennsylvania) for Sperry's line of personal computers. The computers will thus be able to receive telephone calls, handle two-way messages, and accept voice input of data. The Votan card is transparent to most Sperry software, so the keyboard can be totally bypassed without reprogramming. The Votan VPC Voice Card uses



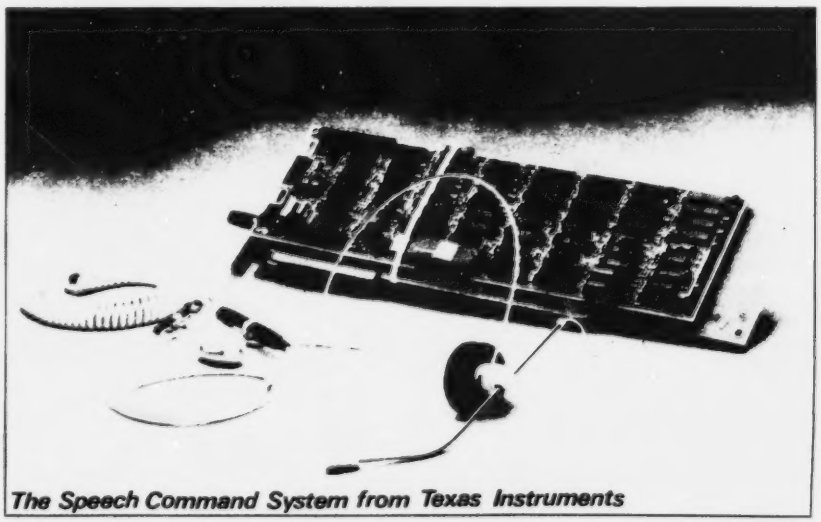
continuous speaker-dependent recognition that will work over telephone lines, permitting access to the computer from any telephone. The telephone management software also provides voice mail facilities and touch tone recognition as well.



The Speech Recognition Keyboard from Key Tronic

There are many other single-board inexpensive products for personal computers. The CallText 5000 from Speech Plus Incorporated (Mountain View, California) is a single-board text-to-speech product for the IBM PC and compatibles. The CallText 5000 can be used in local applications or it can be accessed from any touch-tone telephone. Since this type of system can also originate calls, it can be used to provide a voice version of electronic mail, telexes or mailgrams. The MicroPhonics Technology Corporation's (Renton, Washington) Pronounce system which was just introduced in the fall of 1985 is used in many instances where people can not use a keyboard to enter data -- because of harsh environments, hands-full machinery operation, or physical handicaps. Key Tronic's (Spokane, Washington) Speech Recognition Keyboard and Texas Instrument's (Dallas, Texas) Speech Command System are other popular products.

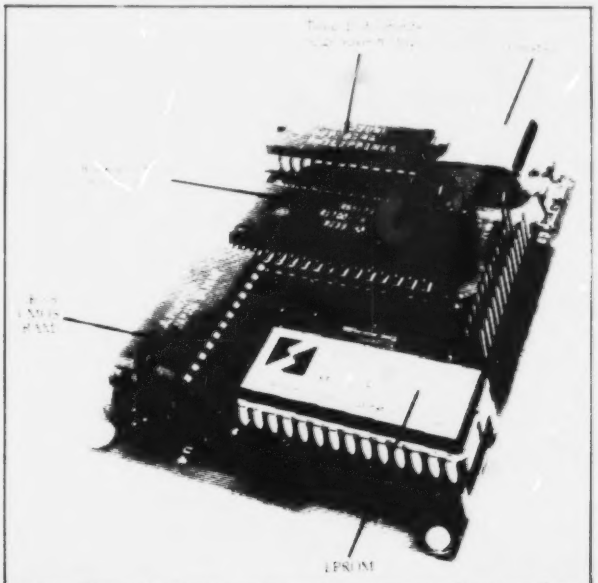
As computing power and memory become cheaper, it will become cost-effective to make voice recognition systems that are much more versatile than today's breed. Less expensive, higher-capacity memory chips will store enough variations of each word to allow speaker-independent operation for thousands of words. Similarly, expected cost reductions in logic circuitry should make feasible the enormously complex processing needed for continuous speech recognition in a large-vocabulary system -- where the system essentially must assume at every instant that a new word might be starting, and proceeded to search for a match. The consensus in the field, however, is that the large-vocabulary, speaker-independent, continuous-speech recognizer will not appear until the mid-1990s (14).



The Speech Command System from Texas Instruments

VOICE SYNTHESIS CAPABILITIES

Until recently, voice output systems could be classified into two broad categories depending on the nature of the synthesis process. Speech-coding systems used a fixed set of parameters to reproduce a relatively limited vocabulary of utterances. These systems produced intelligible and acceptable speech at the cost of flexibility in terms of the range of utterances that could be produced. In contrast, synthetic speech produced by rule provided less intelligible and less natural sounding speech, but these systems had the capability of automatically converting unrestricted text into speech. Over the last few years, significant improvements in text-to-speech systems have begun to eliminate the advantages of simple coded-speech voice response systems over text-to-speech systems. Extensive research on improving the letter-to-sound rules and phonetic implementation rules used by these systems suggest that, in the near future, unrestricted text-to-speech voice response devices may produce highly intelligible and very natural sounding synthetic speech.



Echo Speech Synthesizer from Street Electronics

Two of the several systems available for the IBM Personal Computer include the Echo Speech Synthesizer by Street Electronics Corporation (Capinteria, California) and Speech Master by Tecmar Inc. (Cleveland, Ohio). These single-board systems are based on a fixed list of words: the Speech Master has a 147-word vocabulary while Echo has close to 700 words. The Speech Master uses a Votrax SC01 speech chip and Echo uses the Texas Instruments 5220 speech chip. Both tend to sound somewhat machinelike or robotic.

Several of the newest synthetic speech systems, such as DECtalk, Infovox, the Speech Plus Prose-2000, Votrax Type-N-Talk, and the MITalk-79 research system, provide capabilities for synthesis of different languages and generation of several different synthetic voices.

The standard measure of the intelligibility of synthetic speech is the Modified Rhyme Test in which subjects are required to identify a single word by choosing one of six alternative response words differing by a single phoneme in either the initial or final position. Thus the Modified

Rhyme Test provides a measure of the performance of the system in producing either the initial or final phoneme of a set of spoken words. The test can be used on individual words alone or within a sentence.

The level of performance observed for the DECtalk system comes the closest to natural speech and is considerably higher than performance for any of the other text-to-speech systems studied at the Speech Research Laboratory at Indiana University (15). Performance on MITalk-produced speech was somewhat lower than the DECtalk voices at 93.1 percent correct word identification. The Prose-2000 produced speech that was identified at 87.6 percent correct. The Infovox multilingual system produced English speech that was identified at 87.4 percent correct. Finally, the least intelligible synthetic speech was produced by the Votrax Type-N-Talk with only 67.2 percent correct identification. These results, obtained under closely matched testing conditions in the same laboratory environment, show a wide range of variation among currently available text-to-speech systems.

PERFORMANCE OF SPEECH SYNTHESIS SYSTEMS

System	Percent Correct Word Recognition	
	MRT	WRS
Natural Speech	99.4	99.2
MITalk-79	93.1	93.3
Type-N-Talk	66.2	--
DECtalk	96.7	95.3
Prose-2000	94.3	91.0
Infovox	87.4	--

MRT - Modified Rhyme Test
WRS - Word Recognition in Sentences

6. APPLICATIONS

WHAT YOU CAN DO

Speech synthesis and speech recognition have proven effective in both local and over-the-telephone applications. Local applications include instruction by computer, personal productivity increases (giving your computer verbal rather than typed commands, for instance), alarms, data input by quality-assurance inspectors and data input by shipping and receiving clerks wearing wireless microphones. These are all local applications that leave your hands and eyes free for other work.

Telephone applications are perhaps the more exciting ones. For example, with a good speech synthesis (and sometimes speech recognition) system you can eliminate operators whose main function is to give out specific information repeatedly. This type of application can include giving out status information, bank balances, the location of shipments, information on claim forms, inventory levels on specific items and so on.

Service-based businesses are particularly interested in over-the-phone voice systems because the new technology can deliver more service with a lower payroll demand. Some service-based businesses have installed 24-hour voice systems to handle their customers' routine information needs.

Other applications include voice mail messaging and remote order entry. You can turn any telephone into an interactive data terminal if you connect it to a voice system that understand touch tones. Since everyone knows how to use a telephone, employees need little training to use an over-the-phone data input or delivery system.

WHAT THE DISABLED HAVE DONE

The application of voice technology, or rather of the packages of voice technology, computers, robotics, artificial intelligence and natural language processors, is and will be a key element of efforts to remove the dichotomy between independence and the elderly and handicapped.

In Pittsburgh, 13-year-old Lee Kweller uses a protable computer speech synthesizer that provides him with a voice that speaks and chants English and Hebrew, in near-human tones like those of a 13-year-old boy, even though he has cerebral palsy and is unable to walk or talk. Kweller's new voice was programmed for him by a team led by John Elenberg, the director of Michigan State University's Aritical Language Laboratory (16).

Eugene Spalding is a young man in Columbus, Georgia who was paralyzed from the neck down in a swimming accident. He uses a computer-driven voice-recognition system, for which he has written almost all his own programs, to control various household applicances, stereo, lights, etc., around the house (17). Gene is also developing a self-contained, voice-controlled wheelchair which will incorporate voice recognition technology.

Vitro Laboratories, a company in Silver Spring, Maryland has had a long history of providing special equipment and facilities for the disabled. Tim Forder is an employee who is visually-impaired. Tim uses a talking terminal called Total Talk by Maryland Computer Services (Forest Hill, Maryland) to give him immediate access to data on the company's central computer that he needs to complete reports and to keep records that he is responsible for (17).

David Kostyshyn is a vocational rehabilitation counselor for the Ontario Ministry of Community and Social Services. Since acquiring an Information Thru Speech (ITS) talking computer, he has been able to increase his caseload of handicapped people from 10 to 50 cases. Legally blind, David is continually increasing his efficiency and independence, thanks to the capabilities of his talking computer (18).

The Artificial Intelligence Center of Boeing Computer Services in Seattle, Washington has developed a comprehensive voice-activated microprocessor-based workstation using commercial products such as the IBM PC and Microphonics Pronounce voice recognition system. The workstation allows a physically limited individual to manipulate data electronically and conduct business activities naturally. The voice-activated computer controls a robotic arm that can be used to display reading material, handle filing in a file cabinet, handle floppy diskettes, adjust a printer, and perform miscellaneous support activities (19).

All of these systems require some sort of input and provide a means of obtaining output of one form or another. The type of input that is used

is obviously related to the application considered. For instance, the input to a reading machine for the blind (the Kurzweil optical reader, for example) has to be written material. The crucial component then is an optical character recognizer. It will output characters, which in turn will be used as the input to a speech-to-text voice synthesis system which can be used in many varied applications, such as reading mail over the phone to the visually impaired.

Another example of special input is the Magic Wand Speaking Reader from Texas Instruments. This system uses bar codes to generate synthetic speech from allophones. An optical wand is used to read bar-coded allophone strings from a printed book.

A more interesting type of input is conceptual input, such as a set of Bliss symbols. With this, the user can enter a series of 'concepts' or keywords and the machine will compose a correct sentence from these concepts and sythnesize it. The recent developments of portable speech synthesizers allow for the creation of talking Bliss boards, such as that of the Royal Institute of Technology in Stockholm (20) or the Sahara 2 system at the Centre National d'Etudes des Telecommunications in Lannion, France (21).

But, of course, one of the most interesting inputs is still the human voice coupled with natural language processors and providing output that will control robotic systems or computer control systems.

SOME OPPORTUNITIES IN TRANSPORTATION

The need to have control of our environment and control of our lives, in order to reduce dependence encompasses two types of control. First, it is access to and control of objects in the individuals **immediate** physical surroundings often by using assistive devices. Second, it is access to and control of items **outside** of the individual's immediate surroundings through the applications of communications technology.

For certain types of disabilities and for older persons with declining abilities and who become increasingly less mobile, voice-controlled technologies could (1) become an essential device to ameliorate lost functioning, and (2) become an essential communication link with the outside world to augment lost mobility.

In the first instance, voice-controlled technologies offer the potential to help maintain individuals in their homes longer, and offer increased independence for care givers as well as the older, disabled individual. For the elderly and disabled, it probably offers the most 'natural' approach to controlling objects and devices within their personal space.

In the second instance, voice-controlled technologies could provide a means of improving an increasingly-difficult communication ability with the outside world. This is the primary interest of this report, dealing as it does with the application of voice-controlled technologies to the transportation of the elderly and disabled. The appendices provide an overview of the requirements of various transportation modes in order to make them more accessible for individuals with functional limitations.

From this overview, it is seen that the primary requirements for the elderly and handicapped are related to the exchange of information on the transportation system of concern. Thus, the following concepts are proposed for further investigation in this area:

1. **Computer-Aided Trip Reservation Information.** Voice-controlled trip reservation, trip confirmation and trip status systems for operators of special transportation services for the disabled.

2. **Computer-Aided Travel Information at Terminals.** Voice-synthesis personal information systems for disabled travellers at major transportation terminals such as airports and large railway stations.

3. **Computer-Controlled Personal Wheelchairs.** Voice-controlled wheelchairs offering a self-contained, personal means of transportation without the need for the use of hands or feet.

4. **'Talking' Transportation Devices.** Voice-synthesis systems for multiple transportation system information uses, such as in talking elevators, talking lights, talking bus stops, passenger stop information, etc.

5. **Acceptability of Voice Technologies.** Research on the acceptability of the various voice technologies to the potential elderly and disabled users, their concerns and desires in communicating with machines.

Computer-Aided Trip Reservation Information

In virtually every community in Canada, a personalized form of special transit services is being provided to the elderly and disabled. In order to use these services, the user must register with the operating agency and then phone to reserve trips whenever a trip is desired. Since it is a demand-responsive trip reservation service, it is very management and communications intensive. In an effort to improve efficiency and quality, operators are beginning to adopt computers for use in the scheduling function. The Transportation Development Centre is in fact currently co-sponsoring the demonstration of computerized scheduling systems in both Vancouver and Edmonton.

In conjunction with this current project, it is proposed that the TDC investigate the feasibility of automatic telephone messaging systems for both the trip reservation function and providing trip status information to clients. Automatic telephone callback devices could be connected to the computerized scheduling system to provide automatic confirmation of the client's trip request and also, possibly, to inform the client when the vehicle will arrive and to request that they please make themselves ready. A further study could be undertaken of the feasibility of voice recognition systems in conjunction with a voice synthesis system to actually take clients trip request (this has been attempted in a preliminary fashion for airline reservation systems).

Computer-Aided Travel Information at Terminals

Getting travel information at an airport or bus terminal can pose major problems for the individual who has difficulties hearing, speaking or seeing. Consequently, the Transportation Development Centre is sponsoring the development of a free-standing modular centre that will combine visual and audio messages. An advanced speech synthesizer will allow the partially sighted to get information aurally. The system will also print maps for the location of facilities in the terminal. The prototype communications centre is to be installed at Vancouver International Airport during Expo '86 and will provide information on services in the terminal as well as the usual details that a visitor might need on transportation and accomodations.

In conjunction with this current project, it is proposed that TDC investigate the feasibility of a more personalized form of providing transportation information using voice-synthesis technology. The possible application being proposed is to bond a short length of magnetic tape on the passengers ticket which would be programmed upon arrival at the ticket counter (or at the above communication centre) for the passengers destination and flight number so that it may activate sensors in computer terminals located strategically throughout the airport. Whenever the user is confronted by a decision, insertion of the ticket into the computer terminal will provide personalized voice-synthesized directional information for the user.

Computer-Controlled Personal Wheelchairs

The basic concept of a voice-activated wheelchair is not new and various other institutions are actively investigating this potential application ov voice technology, notably the University of Carleton in Ottawa. This and other robotic aids have been researched with 'control by voice' in mind. There are still a large number of unresolved issues, however, foremost of which is the safety issues encountered.

The Transportation Development Centre has an enviable record in wheelchair research and development and additional research on computer-controlled wheelchairs would appear to fit into the current program. For the disabled user, whose wheelchair must serve as a mobile micro-environment for their compromised physical abilities, the concept of voice-controlled robotic aids is of high interest. A number of attributes still need research that are thought to be desirable (e.g. overcoming noise in the environment, the ease of integrating wheel-chair robotic functions with various voice recognition operations, and the ability of robotic functions to operate under a multi-tasking environment). Future research could be oriented towards an investigation of the feasibility of related speech-controlled knowledge-based expert systems in connection with various sensors and machine vision.

'Talking' Transportation Devices

For elderly and disabled users, voice technology could provide a natural means of communicating with new and unfamiliar (as well as old and familiar) technology. In the transportation area, the need for greater and more timely information for all travellers has been well-documented. The benefits of voice communication to the visually-impaired are obvious, but voice communication can also provide an aid to cognition for a wide spectrum of users with short term memory deficits.

Voice-synthesis systems have the potential of assisting the elderly and disabled in orienting themselves to both the transportation system and the environment in elevators, at pedestrian traffic signals, at transit stops and stations, and on-board transit and transportation vehicles (including the private automobile). Because of the vast potential for applications of voice synthesis technology to transportation of the elderly and disabled, it is proposed that the Transportation Development Centre conduct research in the generalized area of voice synthesis transportation assistive devices.

Acceptability of Voice Technologies

Development of effective applications of voice technology to transportation of the elderly and disabled will depend not only on the willingness of the potential users to use the applications but on the applications ability to communicate with the users and the users ability to understand and comprehend the messages being transmitted.

Research on the acceptability of the various voice technologies to the potential elderly and disabled users, and their concerns and desires in communicating with machines appears to be an area that requires further research. It has been noted, for example, that automatic telephone information systems providing bus schedule information are more likely to

be used by the young and by males. It is proposed therefore that the Transportation Development Centre conduct further research in the area of human factors analysis of new technology. Both the quality and content of voice input/output will be important considerations. The integration and interaction of different input, output and feedback modes should be explored in depth.

Consideration of the capabilities of the elderly and disabled will be critical to the successful application of any type of voice technology. Their cognitive status, their hearing losses, and their visual abilities are three main factors in designing either input control schemes or effective verbal/verbal or verbal/visual feedback. For example, a voice input/output system might query its user "Can you hear this?" and based on the user's response (or non-response) adjust its volume levels to adapt to situations of high ambient noise or declining hearing. These and other factors could probably be most effectively investigated through intensive focus group market research surveys where the voice technology systems could be demonstrated and actively tested.

7. DIRECTIONS

The advantage of voice recognition and response technology is that it allows human beings and machines to communicate through the spoken word. Because people communicate through speech, automatic speech recognition and synthesis provides an interface that accommodates machine users rather than forcing users to conform to a computer's communication constraints. The visually-impaired and elderly can subsequently be significant beneficiaries of this technology.

The obvious benefits can be extended by natural-language processing to achieve a highly adapted human-to-machine interface. When the machine of interaction is a computer with a degree of artifical intelligence and a means of control of robotic aids, assistive devices and machine vision, the advantages for the elderly and disabled in controlling their environment would appear to be without limit. The realization of some of these benefits of voice technology would contribute significantly to an increase in independence for the elderly and disabled, enhancing their sense of freedom. As voice technology moves out of the laboratories and into widespread use, its potential advantages to persons with a range of disabilities will become even more readily apparent.

The first major application of voice technology is in access to and control of objects in the individuals immediate physical surroundings by using other assistive devices controlled by the voice units. The second major application, the one of primary interest in this report, is access to and control of items outside of the individual's immediate surroundings through the application of telecommunication systems. A brief review of the requirements for improved transportation services for the elderly and disabled reveals a preponderance of requirements for information.

DIRECTIONS

Various applications of voice technology to transportation for the elderly and disabled were identified in the previous chapter: telephone voice response systems, personal transportation terminal directional systems, voice-controlled wheelchairs, and talking transportation assistive devices; as well as a need for more research on the acceptability of the technology to the end user.

Clearly, the next step would be to evaluate the technical, operational, and economic feasibility of the potential applications -- either through further intensive research or through selected demonstration projects (22). The following is intended to provide directions for this research effort and can be considered as research statements for each application.

DIRECTIONS FOR ASSESSING VOICE TECHNOLOGY APPLICATIONS

Technical Feasibility

To determine the technical feasibility of using voice technology, the evaluation team must estimate minimal and optimal voice input/output performance requirements by determining the vocabulary size and grammatical complexity required, accuracy needed, effects of inaccuracy, health and safety concerns, and environmental and human factors inherent in the application.

Vocabulary Complexity. The grammatical complexity required for an application may range from single word commands to specified responses to freely-composed paragraphs. The technical capabilities required of the voice technology system cannot be determined until application requirements are thoroughly understood.

Accuracy. Voice recognition and response accuracy will probably be less than 100 percent. Accuracy requirements must be evaluated in terms of the consequences of an error. The evaluation team must ask how critical accuracy is in the application and how errors are detected and corrected by the voice system being considered.

Environmental Factors. Environmental factors also affect the technical feasibility of a voice technology application and should be thoroughly investigated. The type and level of ambient and transient noise are the primary environmental factors; others that should be considered are temperature, humidity and air quality.

Human Factors. Perhaps the most important factor in determining feasibility is the human factor since the application has to be designed to fit the user, not the other way around. Perhaps the first issue to determine is whether the user is currently comfortable in a voice environment.

Operational Feasibility

The operational feasibility of voice technology is largely a question of user acceptance of the new technology: how will the proposed modifications of existing systems be received by those directly affected.

Apart from a natural resistance to change, user acceptance of voice technology may be determined by the following: perceived potential benefits and appropriateness of the technology, presentation and explanation of the new technology, and method of implementation.

Benefits and Appropriateness. Some applications are simply not viewed by users as being appropriate for voice technologies, or they do not see the benefit of adding voice communication capabilities. Talking vending machines, for example, are considered inappropriate by the general public and this feature has been withdrawn from many such machines. On the other hand, initial resistance to voice recognition has been overcome in many business applications where its labour-saving value has been convincingly demonstrated.

Presentation and Explanation. User acceptance of voice technology will also depend significantly on the presentation and explanation of the new technology. Does the application appear as a simplification or an unnecessary complication of the existing way of doing things? Is the technology presented as a cost-saving step for the transportation operator or as a life-enhancement device for the user?

Method of Implementation. A final consideration in terms of user acceptance is the way the new technology is implemented. Will the old method continue to be used as a backup or alternative method? If it is, the new technology may be consistently ignored in favour of the established method.

Economic Feasibility

Evaluating the economic feasiblity of implementing voice technology hinges on determining what benefits can be expected. Savings may be derived from increased productivity for transportation operators, or from mobility and life enhancement for the users.

Cost estimates of implementing voice technology should include not only the cost of equipment, but the costs of interfacing equipment to the existing system, both in terms of hardware and software. These costs can result in a final cost that far exceeds the cost of voice technology itself. Systems can range from a few hundred dollars for a board or chip to over $60,000 for a stand-alone system.

RECOMMENDED COURSE OF ACTION

It is recommended that the Transportation Development Centre initiate a research program oriented towards the application of voice technology to the transportation needs of the elderly and disabled. Several potential applications have been identified in this report and there are undoubtedly many more. The problem of acceptance by individual users will have to be addressed. The general resistance to new technology in any application must be overcome as well as specific problems. User acceptance is affected by the anticipated benefits and methods of presentation and implementation, as well as device reliability, recognition robustness, and naturalness of synthesized speech. Technological advances are expected to address these later problems so that use of voice technology will become increasingly comfortable for users. A research program is necessary to address some of the issues raised by the other problems. Finally, applications should be based on documented and demonstrated performance on standard feasibility tests and a thorough understanding of the effort required to implement the application, expecially when voice will significantly modify the way users have accessed transportation systems and services in the past.

REFERENCES

1. Aarons, Dick. "The Voice of the '80's". PC MAGAZINE, Volume 4, Number 5, March 1985. pp 114-121.

2. "Speech Technology - A Management Overview". DataPro Research Corporation Publication No. TC41-007E-201. 1983.

3. "Voice Store-and-Forward Systems". DataPro Research Corporation Publication No. TC41-007E-101. 1984.

4. Terrill, Eileen T. "Listening with the Brain: Human Speech Perception". INCIDER MAGAZINE, July 1985. pg 16.

5. Schoen, John. "When You Talk, Your PC Listens". PC MAGAZINE, Volume 4, Number 5, March 1985. pp 122-132.

6. Alvernaz, B. "If I Only Had a Voice...". PC MAGAZINE, Volume 2, Number 4, September 1983. pp 511-524.

7. Atal, B. S. "Linear Predictive Coding of Speech". In COMPUTER SPEECH PROCESSING, Prentice-Hall International, 1985. pp 81-124.

8. Veltri, Steven J. HOW TO MAKE YOUR COMPUTER TALK. McGraw-Hill Book Company. 1985. 256p.

9. Allen, Jonathan. "A Perspective on Man-Machine Communication by Speech". PROCEEDINGS OF THE IEEE, Volume 73, Number 11, November 1985. pp 1541-1551.

10. Schalk, Thomas B. "Speech Recognition: Machines That Listen". ROBOTICS AGE, Volume 7, Number 4, April 1985. pp 11-14.

11. Poole, Harry H. "Voice Recognition". Auerbach Publishers Inc. Publication No. 004.0001.011. undated.

12. Dickson, Craig R. A FLEXIBLE SYNTHETIC SPEECH DELIVERY SYSTEM. Technical Report No. SS8401, Centre for Technology Research, Victoria, B.C. February 1984. 10p.

13. Lazzaor, J. J. "Taking Instead of Typing: Computers are Gaining the Ability to Recognize and Respond to Voices". In HIGH TECHNOLOGY, Volume 6, Number 1, January 1986. pp 58-59.

14. Rosenberg, Robert. "Speech Technology Takes Slow But Steady Commercial Course". ELECTRONICS WEEK, April 22, 1985. pp 34-35.

15. Pisoni, D. B. et al. "Perception of Synthetic Speech Generated by Rule". In PROCEEDINGS OF THE IEEE, Volume 73, Number 11, November 1985. pp 1665-1676.

16. "The Gift of Voice Synthesis". NEWSWEEK MAGAZINE, November 4, 1985. pp 65-66

17. Scott, Victoria. "Life Enhancement for the Handicapped through Voice Technology". SPEECH TECHNOLOGY, April 1982. pp 63-65

18. Williams, John M. "Speech Output Helps Disabled People Lead Productive Lives". SPEECH TECHNOLOGY, Sept/Oct 1982. pp 69-72

19. Fu, Caroline. "Voice Driven Robotic Aids for the Physically Limited". Proceedings of the Aid to the Disabled Session of the Voice I/O Applications Conference '85, Sar Francisco, California. September 1985.

20. Carlson, R. et al. "Bliss Communication with Speech or Text Output". IEEE ICASSP, 1982. pp 747-750.

21. Stella, M. "Speech Synthesis". In COMPUTER SPEECH PROCESSING, Prentice-Hall International, 1985. pp 421-460.

22. Waltrous, Raymond. "Outlook on Speech I/O Technology". Auerbach Pulishers Publication No. 53-10-09, Data Communications Management, 1985. 10p.

APPENDICES

Acoustic Phonetics. The physical characteristics of a speech signal.

Algorithm. A set of instructions to solve a particular problem, such as finding a square root, or recognizing speech.

Allophone. A variation of a particular phoneme sound that depends on its position in a word.

Amplitude. In waveforms, the range from the zero axis (mean) to the maximum value (extreme). In sound waves, the amplitude can be regarded as the volume of the signal.

Analog Filter. A device that is used to screen out or pass through analog waveforms.

Analog-to-Digital Conversion. The process of converting an analog waveform to a digital waveform.

Analog Waveform. A waveform that is continuous.

Articulation. The process of positioning or connecting the articulators and changing the shape of the vocal tract when producing speech.

Articulator. A medium by which sounds are created. In the human vocal tract, the articulators are the lips, tongue, palate, and teeth, and they shape the human oral cavity.

Articulatory Phonetics. The process of making speech sounds or phonemes.

Coarticulation. The blending together of two allophones.

CSR. Continuous-speech recognition, the identification of spoken words without unnaturally-long pauses between them.

Digital-to-Analog Conversion. The process of converting a digital waveform to an analog waveform.

Digital Waveform. A waveform that is used to represent an analog waveform using digital values.

Discrete Logic. Transistor transistor logic (TTL) which comprises electronic gates.

DSUR. Discrete-utterance speech recognition -- single words with distinct pauses between them.

Dynamics. In human speech, the process of varying pitch in any utterance.

Formants. The resonant frequencies of the vocal tract.

Intonation. The manner of producing tones with regard to accurate pitch.

LPC. Linear predictive coding, a mathematical algorithm predicting the next most likely event based on the previous data.

Phonemes. The smallest unit of speech that is used to distinguish meanings between words.

Pitch. The highness or lowness of a sound due to vibrations of sound waves.

Resonant Frequency. Any particular sound is measured by the number of times the sound wave oscillates in a given period. A sound frequency is measured in cycles per second or hertz (Hz). For example, the sound frequency of human speech ranges from 0-5,000 cycles per second or 0-5,000 Hz.

Sampling Frequency. The rate at which an analog waveform is 'looked at' when converting to a digital waveform.

Speaker-dependent. A voice-recognition system that must be initialized by the person who will use it.

Speaker-independent. A voice-recognition system that can respond to any speaker and requires no initialization by the end user.

Speech Digitization. Conversion of the conventional analog voice waveform into a digital signal to allow computer processing of the information or to improve communications security or reliability.

Speech Recognition. The ability of a system to recognize certain words from a fixed (usually small) vocabulary, allowing the computer to take some action based on this recognition.

Speech Synthesis. Artificial production of speech through a computer or a special integrated circuit chip, using one of the several types of speech generation available:

-Formant synthesis. Based on electronic modeling of the natural resonances of the human vocal tract.
-Phoneme synthesis. Uses special parameters derived from the basic units of speech (phonemes).
-Linear predictive coding (LPC). Based on certain known characteristics of human speech; uses the repetitive nature of most speech sounds to reduce the number of bits necessary for speech fidelity.
-Waveform synthesis. Uses the stored digitized amplitude characteristics of the speech waveform.

Template. A disk file containing words or phrases to be used as the basis for speech comparison in a voice-recognition system.

Trachea. The throat or windpipe.

Unvoiced Sounds. Sounds that are produced in the vocal tract when the vocal cords are not vibrating.

Vocal Tract. In humans, it mainly consists of the articulators, oral cavity, nasal cavity, and throat.

Voice Store-and-Forward. As in data store-and-forward techniques, storing voice signals (usually in digital form) and then transmitting them to the desired recipient when a channel becomes available, when a connection is made, or at a particular time of day.

Voiced Sounds. Sounds that are produced in the vocal tract when the vocal cords are vibrating.

American Microsystems Inc., 3800 Homestead Rd., Santa Clara, California 95051 (408) 246-0330 **2**

Amnet Corporation, 101 Morse Street, Watertown, Massachusetts 02172 (617) 923-1850 **3**

Artec Systems, 9104 Red Branch Road, Columbia, Maryland 21045 (301) 730-1237 **2**

Artic Technologies, 2234 Star Court, Auburn Heights, Michigan 48057 (313) 852-8344 **2**

AT&T Information Systems, One Speedwell Avenue, Headquarters Plaza West Tower, Morristown, NJ 07060 (201) 898-2000 **3**

AudoPilot Incorporated, P.O. Box 3006, 516 Walt Whitman Rd., Huntington Station, New York 11746 (516) 351-4862 **1**

Auricle Incorporated, 20823 Stevens Creek Blvd., Cupertino, California 95014 (408) 257-9830 **1**

BBL Industries Incorporated, P.O. Box 48488, Atlanta, Georgia 30362 (404) 449-7740 **3**

Centigram Corporation, 155A Moffett Park Drive, Sunnyvale, California 94086 (408) 734-3222 **1,2,3**

Cognitronics Corporation, 25 Crescent Street, Stamford, Connecticut 06906 (203) 327-5307 **3**

Commterm Incorporated, Ten Third Avenue, Burlington, Massachusetts 01803 (617) 663-4442 **3**

Computer Consoles Inc., 97 Humboldt Street, Rochester, New York 14609 (716) 482-5000 **3**

Computer Curriculum Corp., 1070 Arastradero Rd., Box 10080, Palo Alto, California 94303 (415) 494-8450 **3**

Dialogic Corporation, 164 McKinley Ave., East Hanover, New Jersey 07936 (201) 386-0202 **1**

Digital Equipment Corporation, 200 Baker Avenue, West Concord, Massachusetts 01742 (617) 493-TALK **2**

Digital Pathways Inc., 1260 L'Avenida, Mountain View, California 94043 (415) 969-7600 **3**

Dragon Systems Inc. (OEM), 173 Highland Street, West Newton, Massachusetts 02165 (617) 527-0372 **1,2**

Engineered Systems Inc., 14775 Grover Street, Omaha, Nebraska 68144 (402) 333-0100 **3**

Ericsson Information Systems, 7465 Lampson Avenue, Garden Grove, California 92641 (714) 635-0150 **3**

E-Systems Incorporated, Garland Division, P.O. Box 660023, Dallas, Texas 75266-0023 **2**

General Instruments, 600 West John Street, Hicksville, New York 11802 (516) 733-3000 **2**

GTE Business Communication Systems, 12502 Sunrise Valley Drive, P.O. Box 4050, Reston, Virginia 22090 (703) 435-7643 **3**

Harris Semiconductor Inc., P.O. Box 883, M/S 58-36, Melbourne, Florida 32901 (305) 724-7000 **2**

IBM Corporation, 1133 Westchester Avenue, White Plains, New York 10604 (914) 696-1900 **3**

Infolink Data Communications, 1925 Holste, Northbrook, Illinois 60062 (312) 291-2900 **3**

Interface Systems International, P.O. Box 3289, Portland, Oregon 97208 (503) 256-3214 **2**

Interface Technology, 10500 Kahlmeter Drive, St. Louis, Missouri 63132 (314) 426-6880 **3**

Interstate Electronics Corp., 1001 E. Ball Avenue, Box 3117, Anaheim, California 92803 (714) 772-2811 **1,3**

Interstate Voice Products, 1849 West Sequoia Avenue, Orange, California 92668 (714) 937-9010 **1**

Keytronic Corporation, P.O. Box 14687, Spokane, Washington 99214 (509) 928-8000 **1**

Kurzweil Applied Intelligence Inc., Suite 1100, 2049 Century Park East, Los Angeles, California 90067 (213) 203-9909 1

Maryland Computer Services Inc., 2010 Rock Spring Avenue, Forest Hill, Maryland 21050 (301) 879-3366 2

Micromint Incorporated, 25 Terrace Drive, Vernon, Connecticut 06066 1

Microphonics Technology Corp., 234 S.W. 43rd Street, P.O. Box 7458, Renton, Washington 98057 (206) 251-9009 1

Mimic Electronics Company, 1 Mohegan Road, Acton, Massachusetts 01720 (617) 263-5837 2

National Semiconductor Corp., 1900 Semiconductor Drive, Santa Clara, California 95051 (408) 721-5000 2

Natural Microsystems Corp., Natick, Massachusetts (617) 655-0700, (617) 651-2186 for demo 2

NEC America Incorporated, 8 Old Sod Farm Road, Melville, New York 11747 (516) 753-7000, Michiyuki Uenohara, Vice President 1

Northern Telecom Incorporated, 2305 Mission College Boulevard, Santa Clara, California 95050 (408) 988-5550 3

Octel Communications Incorporated, 1841 Zanker Road, San Jose, California 95112 (408) 947-4500 3

Optimum Communications Incorporated, 538 Oakmead Parkway, Sunnyvale, California 94086 (408) 749-1911 3

Panasonic Corporation, One Panasonic Way, Secaucus, New Jersey 07094 (201) 348-5270 2

Perception Technology Corp., 50 Shawnut Road, Canton, Massachusetts 02021 (617) 821-0320 3

Periphonics Corporation, 25 Orville Drive, Bohemia, New York 11716 (516) 467-0500 3

Rolm Corporation, 4900 Old Ironside Drive, Santa Clara, California 95050 (408) 988-2900 3

Scott Instruments, 1111 Willow Springs Drive, Denton, Texas 76205 (817) 387-9514 1

Solid State Systems, 1900 Delk Industrial Boulevard, Marietta, Georgia 300067 (404) 952-9401 **3**

Sperry Corporation, P.O. Box 500, Blue Bell, Pennsylvania 19424 (215) 542-4011 **3**

Street Electronics Corporation, 1140 Mark Avenue, Carpinteria, California 93013 (805) 684-4593 **2**

Sun Electronic Corporation, 525 West Remington Drive, Sunnyvale, California 94086 (408) 245-3511 **1**

Tecmar Incorporated, 6625 Cochran Road, Cleveland, Ohio 44139-3377 (216) 349-0600 **1,2**

Telesensory Systems Inc., 3408 Hillview Avenue, Palo Alto, California 94304 (415) 493-2626 **3**

Texas Instruments Inc., Data Systems Group, P.O. Box 402430, Dallas, Texas 75240 (214) 994-6611 **1**

Threshold Technology, 1829 Underwood Boulevard, Delran, New Jersey 08075 (609) 461-9200 **1**

Triformation Systems Inc., 3132 S.E. Jay Street, Stuart, Florida 33494 (305) 283-4817 **2**

Verbex Corporation, 2 Oak Park, Bedford, Massachusetts 01730 (617) 275-5160 **1**

VMX Incorporated, 1241 Columbia Drive, Richardson, Texas 75081 (214) 699-1461 **3**

Voicemail International Inc., 2225 Martin Avenue, Santa Clara, California (408) 496-6555 **3**

Voicetek Corporation, 61 Chapel Street, Newton, Massachusetts 02195 (617) 964-8820 **1**

Voice-Works Incorporated, 74 Alex Avenue, Woodbridge, Ontario L4L 4K6 (416) 851-7751 **2**

Votan Corporation, 4487 Technology Drive, Fremont, California 94538 (415) 490-7600, (415) 490-7977 for demo **1,3**

Votrax Division, Federal Screw Works, 500 Stephenson Highway, Troy, Mississippi 48084 (313) 588-2050 **2,3**

Vynet Corporation, 2607 South Winchester Boulevard, Campbell, California 95131 (408) 370-9764, (408) 349-0600 for demo **2,3**

Wang Laboratories, 1 Industrial Avenue, Lowell, Massachusetts 01851 (617) 459-5000 **3**

NOTE: At the end of each listing is one or more numbers, in boldface, that correspond to the vendor's area(s) of concentration. The numbers are (1) recognition, (2) synthesis, and (3) compression.

S. E. Levinson, Acoustics Research Department, **AT&T Bell Laboratories,** Murray Hill, New Jersey 07974

Aaron E. Rosenberg, **Bell Laboratories Murray Hill,** 600 Mountain Avenue, Murray Hill, New Jersey 07974 (201) 582-4985

Prof. William Cameron, Neil Squires Foundation, TRIUMF, **University of British Columia,** 2075 Westbrook Mall, Vancouver, B.C. V6T 2A3 (604) 228-4711

P. Ladefoged, Phonetics Laboratory, Linguistics Department, **University of California at Los Angeles,** Los Angeles, California 90024

M. Stella, **Centre National d'Etudes des Telecommunications,** 22301 Lannion, France

Rex N. Dixon, **IBM Thomas J. Watson Research Center,** P.O. Box 216, Yorktown Heights, New York 10598 (914) 945-1780

Dr. David B. Pisoni, Speech Research Laboratory, Department of Psychology, **Indiana University,** Bloomington, Indiana 47405 (812) 335-1155

J. S. Bridel, **Joint Speech Research Unit,** Princess Elizabeth Way, Cheltenham, Gloucestershire, GL52 5AJ, United Kingdom

Robert J. McAulay, **Massachusetts Institute of Technology,** Lincoln Laboratory, 244 Wood Street, P.O. Box 73, Lexington, Massachusetts 02173

L. K. Tyler, **Max Planck Institute for Psycholinguistics,** Nijmegen, The Netherlands

John Eulenberg, Director, Artificial Language Laboratory, Computer Science Department, **Michigan State University,** East Lansing, Michigan 48824 (517) 353-0870

Melvin Hunt, **National Research Council,** National Aeronautics, Building U66, Ottawa, Ontario K1A 0R6 (613) 998-3408

J. R. Duan, **North American Phillips Corporation,** Phillips Laboratories, 354 Scarborough Road, Briarcliff Manor, New York 10510 (914) 762-0300 ext 252

Royal Institute of Technology, Department of Speech Communication, S-100 44 Stockholm, Sweden

E. Deviney, **Siemens Corporation**, Corporate Research adn Technology, 20 Olney Avenue, P.O. Box 5476, Cherry Hill, New Jersey 08034 (609) 424-2856

B. Craig Dickson, Research Director, Centre for Speech Technology, **University of Victoria**, P.O. Box 1700, B.C. V8W 2Y2 (604) 721-7426

The accompanying comparison charts summarize the characteristics of over a hundred speech products from about fifty vendors. The charts have been divided into four distinct areas: voice recognition systems, voice synthesis systems, voice response systems, and voice messaging systems. The type of system is indicated at the top of each page. The chart entries and their significance are explained in the following paragraphs.

Voice Recognition. The charts start off by providing the purchase price of the system and typical applications. Then the processing facilities are described , including the type of CPU, the amount and type of main memory, and the product's software characteristics. Frequently, the software used is proprietary, and is usually built into the system as firmware. The key characteristics of a voice recognition system that are identified include the recognition technique, the number of speakers and voice references stored, if the system handles continuous speech, in addition to isolated-word recognition. Since the system's performance is a key factor, the average response time (from spoken word to system response), percent recognition accuracy and amount of memory per word is indicated.

Voice Synthesis. These charts are similar to voice recognition, except in the characteristics area. Here the synthesis technique, the maximum vocabulary size (which can be the form of words or total seconds of speech), support of foreign languages, controls supported (such as pitch, tone, volume, and speech rate) are indicated.

Voice Response. The voice reponse systems are identified by the type of system and its configuration, such as a standalone product or a component within a larger system. The maximum number of communication lines that connect the system to a counterpart, such as a PABX, and the type of remote terminals supported, such as touch-tone or CRT, are indicated. Details of the processing facilities are similar to the previous charts, and the methods for interacting with the system are added, as well as the call handling capacity, and type of message generated.

Voice Messaging. The system configuration is identified, the maximum number of communication lines, the type of remote terminals supported, the processing facilities and the methods for interacting with the system. Speech characteristics include the type of message generated, the size, the call handling capacity and possible applications.

VOICE RECOGNITION SYSTEMS

MANUFACTURER AND MODEL	AudoPilot 1000	Artec Systems Micro-Ear	Auricle Inc Auricle 1	Bridgeport BTSR-1
GENERAL INFORMATION				
Purchase Price	$459	$579	$2,480	$2,000
Typical Applications	Hobbyists, education, program data entry and control	Hobbyists, education, toys, handicapped	Development system for OEMs, end user, manufacturing	Development system for OEM's, manufacturing, handicapped
Number of Units Sold	na	na	>1000	na
PROCESSING FACILITIES				
System Processor	IBM PC host	IBM PC, Apple II host	Z-80	na
System Memory	Uses host CPU	16K RAM, ROM	16K RAM	na
System Software	Disk	Proprietary, Disk	Firmware	na
Communications Interface	Plug-in board	RS-232C	RS-232C & IEEE 488	na
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Pattern generation, pattern recognition	Pattern generation, pattern recognition	Spectral analysis, filter amplitudes, pattern recognition	Spectral analysis, total energy, sonorant energy, hf energy, spectral tilt
Speaker Independence	No	No	Yes	No
Handles Continuous Speech	No	No	No, isolated words	No, isolated words
Average Response Time	na	Less than 1 sec.	180 ms	200 ms
Training Passes Required	na	na	3	Monomic
Vocabulary Size (Words)	32/file	256	89-149	20-39
Additional Characteristics	System board for IBM PC - requires 256K and one disk drive	Self-contained system for IBM PC, Apple II	--	--
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	na	na	250-5,000	150-7,000
Number of Filters	na	na	16	22
Bits Per Word	6K/file	na	800	3,200/syllable
Percent Recognition Accuracy	na	95	97	99.2

MANUFACTURER AND MODEL	Centigram Corp MIKE	Dialogic Corp Dialogic 2	Interstate Electronics VRC008/100-1	Interstate Electronics VRM041/102
GENERAL INFORMATION				
Purchase Price	$1,900 to $3,850	$495	Contact vendor	Contact vendor
Typical Applications	Data entry, control functions, file enquiries	Data entry system for IBM PC computer	Data entry, inventory control, consumer systems, toys	Data entry, inventory control, consumer systems, toys
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Z-80	IBM PC host	na	na
System Memory	48K RAM, 4K EPROM	Uses host CPU	External RAM	RAM, ROM
System Software	Proprietary	na	Firmware	Firmware
Communications Interface	RS-232C, parallel	na	Parallel	Parallel, RS-232C
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Spectral analysis, word framing, pattern generation	Pattern generation, pattern recognition, voice compression	Voice pattern waveform analysis, template matching	Pattern generation pattern matching
Speaker Independence	No	Yes	No	No
Handles Continuous Speech	Yes, phrases	No	No	No, isolated words
Average Response Time	na	na	200 ms/250 ms	65 ms/125 ms
Training Passes Required	na	na	na	3-5
Vocabulary Size (Words)	96 words	Unlimited	8/100	40/100
Additional Characteristics	Voice response capability	For IBM PC with 256K RAM and disk drive, incl phone interface	VRC008 is speaker independent, both are chip sets	Both board level systems
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	na	na	na	150-7,000
Number of Filters	na	na	na	16
Bits Per Word	na	na	na	512
Percent Recognition Accuracy	na	na	90/99+	99+

VOICE RECOGNITION SYSTEMS

MANUFACTURER AND MODEL	Interstate Electronics VRT101	Interstate Electronics VRQ400	Interstate Voice Prod. VocaLink	Keytronic Corp. KB 5152V
GENERAL INFORMATION				
Purchase Price	$5,295	na	$1,650	na
Typical Applications	Data entry, inventory control	Data entry, inventory control	Parts inspection, OEM's, inventory control, reservation order processing	Handicapped, quality control, parcel-handling, brokerage houses
Number of Units Sold	750	na	na	na
PROCESSING FACILITIES				
System Processor	Dual Z-80	na	Intel	Motorola 6803
System Memory	48K RAM	RAM, ROM	na	16K RAM, ROM
System Software	VOICE (proprietary)	Firmware	na	Proprietary
Communications Interface	Two RS-232C ports	DEC Q-Bus	25-pin connector	DIN-plug keyboard
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Pattern generation, pattern matching, filter amplitudes	Pattern generation, pattern matching	Pattern generation and matching, delta thresholds	Pattern generation, pattern matching, delta thresholds
Speaker Independence	No, one speaker	No	No	No
Handles Continuous Speech	No, isolated words	No	No, discrete words	No, discrete words
Average Response Time	125 ms	125 ms	na	na
Training Passes Required	3-5	na	na	na
Vocabulary Size (Words)	40-89	100	240	160
Additional Characteristics	Fully self-contained intelligent terminal with voice recognition	Single-board unit for DEC VT103 and LSI-11	Single-board unit for IBM PC - requires 64K and disk drive	Plug-compatible keyboard with microphone for IBM PC
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	150-7,000	na	na	250-7,500
Number of Filters	16	na	16	16
Bits Per Word	na	na	na	128
Percent Recognition Accuracy	97.1	99+	85	na

MANUFACTURER AND MODEL	Kurzweil Appl Int'gence KVS-3000	Micromint Lis'ner 1000	Microphonics Technology Pronounce	NEC SR-10 Voice Plus
GENERAL INFORMATION				
Purchase Price	$6,000	$189	$595	na
Typical Applications	Data entry, control functions, development system for OEMs	Hobbyists, education, handicapped,	Data entry, inventory control, consumer systems,	Data entry, inventory control and robotics, product inspection
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	IBM PC host	Apple II host	IBM PC host	Custom LSI chip
System Memory	na	na	na	RAM, ROM
System Software	Proprietary	Disk	Proprietary	Proprietary
Communications Interface	RS-232C	Plug-in board	na	na
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Accoustic pattern analysis	Spectral analysis, pattern recognition using GI SP-1000 chip	Voice pattern waveform analysis, template matching	Pattern generation pattern matching
Speaker Independence	Yes	No	No	No
Handles Continuous Speech	Yes, phrases	No	No	No
Average Response Time	na	na	na	na
Training Passes Required	na	na	1-2	na
Vocabulary Size (Words)	1,000 words	64 words	128	250
Additional Characteristics	--	Voice response capability with SSI-263 voice chip	Board-compatible system for IBM PC with 256K RAM	Board-compatible system for IBM PC
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	na	na	na	na
Number of Filters	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	95	na	na

VOICE RECOGNITION SYSTEMS

MANUFACTURER AND MODEL	NEC DP-200 CSR	NEC SR-100	NEC SR-1000	NEC SR-2000
GENERAL INFORMATION				
Purchase Price	$20,000	$1,500	$60,000 and up	na
Typical Applications	Baggage sorting, data entry, inventory control, military	Data entry, manufacturing, handicapped	Telephone specialty systems	na na
Number of Units Sold	40	na	>10	na
PROCESSING FACILITIES				
System Processor	NEC DP-200 CSR	na	na	na
System Memory	64K D-RAM	ROM	na	na
System Software	na	na	na	na
Communications Interface	RS-232C, RS-422	RS-232C	na	na
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Dynamic programming	Dynamic programming	na na na	na na na
Speaker Independence	No	No	na	Speaker independent
Handles Continuous Speech	Yes	Yes, connected words	No, isolated words	Continuous speech
Average Response Time	300 ms or continuous	300 ms	na	na
Training Passes Required	1 isol to 4 cont	1 or 2	None	na
Vocabulary Size (Words)	500 isol & 50-150 cont	120	10-19	na
Additional Characteristics	Audio response	--	Suitable for telephone input	New product
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	120-6,000	120-6,000	300-3,000	na
Number of Filters	16	16	na	na
Bits Per Word	4,000	na	na	na
Percent Recognition Accuracy	98.8	99	na	na

MANUFACTURER AND MODEL	Scott Instruments VET-2	Speech Systems VoiceLine	Sun Electronics VR-1500	Tecmar Inc PC Mate Voice Rec
GENERAL INFORMATION				
Purchase Price	$995	$2,480	Contact vendor	$995
Typical Applications	Data entry, inventory, numerical/quality control, hobbyists, handicapped	Development system for OEM's	Data entry, sortation, games, handicapped	Data entry, inventory control, consumer systems, toys
Number of Units Sold	>100	na	na	na
PROCESSING FACILITIES				
System Processor	Apple II/TRS-80 host	Z-80	Z-80	IBM PC host
System Memory	32-48K RAM (host)	16K RAM	16K RAM; 4K ROM	8/16K RAM, ROM
System Software	na	Firmware	na	Proprietary
Communications Interface	RS-232C via host	RS232C & IEEE488	Parallel	RS232C
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Pattern generation, comparison with stored vocabulary	Spectral analysis, pattern recognition	Pattern generation, pattern matching, data compression	Pattern generation pattern matching
Speaker Independence	No	Yes	No	No
Handles Continuous Speech	No	No	No	No
Average Response Time	250 ms	na	400 ms	na
Training Passes Required	7	na	na	na
Vocabulary Size (Words)	40 words	5,000	32	100/200
Additional Characteristics	--	--	--	Board system for IBM with 64K RAM and disk drive
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	250-5,000	na	na	na
Number of Filters	4	na	na	na
Bits Per Word	1,000	na	na	na
Percent Recognition Accuracy	88.4	95	na	95

VOICE RECOGNITION SYSTEMS

MANUFACTURER AND MODEL	Texas Instruments Speech Command	Threshold Technology 500/580/600/680	Threshold Technology VNC-300	Verbex Corp Model 1800
GENERAL INFORMATION				
Purchase Price	$2,600	$8,700-11,700	$42,650-68,765	$65,000 and up
Typical Applications	Computer program control and data entry, games	Speech recognition subsystem for OEMs, military, handicapped	Single-user voice numerical control system	Data entry, banking order processing, airline reservation
Number of Units Sold	na	500/800	46	>10
PROCESSING FACILITIES				
System Processor	TI PC host	LSI-I/LSI-II	DG Nova 4 Eclipse	DEC PDP-11/04
System Memory	Uses host CPU	64K RAM	64K RAM	256K RAM
System Software	Disk	Firmware	Firmware	Proprietary
Communications Interface	Plug-in board	RS-232C	RS-232C	RS-232C, 3270
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Linear predictive coding	Spectral analysis, pattern recognition	Spectral analysis, pattern recognition	Vector processor, multiple reference templates
Speaker Independence	No	Yes	Yes	Yes (8 simultaneously)
Handles Continuous Speech	Yes	No, isolated word	No	Yes
Average Response Time	na	Less than 1 sec.	Less than 1 sec.	300 ms or continuous
Training Passes Required	na	10	na	None
Vocabulary Size (Words)	50	250/340	Unlimited	10-39
Additional Characteristics	Board system for Texas Instruments Professional Computer	Speech recognition subsystem and voice data entry terminal	Single-user voice numerical control system	Optional voice response capability, suitable for telephone input
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	na	150-7,000	na	300-3,000
Number of Filters	na	16	na	32
Bits Per Word	300 sec.	1,000	na	4,000
Percent Recognition Accuracy	na	99	99	99

MANUFACTURER AND MODEL	Voicetek Corp V10-1000 Series	Voice-Works VoiceLink	Votan Corp V10000	Votan Corp VPC 2000 Voice Card
GENERAL INFORMATION				
Purchase Price	$295	na	$3,200-4,200	$2,450
Typical Applications	Data entry, games, education, handicapped	na	Data entry/retrieval, control function, OEMs, military, airborne	Data entry, inventory control, consumer systems, telephone
Number of Units Sold	Over 11,000	na	>100	na
PROCESSING FACILITIES				
System Processor	Microcomputer host	IBM PC host	Digital signal processing	IBM/Sperry PC host
System Memory	na	na	128K RAM	RAM, ROM
System Software	na	na	na	Voice Key software
Communications Interface	IEEE-488, I/O port of various micros	na	RS-232C	Parallel, RS232C
SPEECH CHARACTERISTICS				
Recognition Technique(s)	Pattern generation, pattern matching	na	Dynamic programming, pattern matching, digital filters	Pattern generation pattern matching
Speaker Independence	No	Yes	Yes	No
Handles Continuous Speech	No	No	No	Yes
Average Response Time	0.3 sec.	na	0.3 sec.	na
Training Passes Required	na	na	2	2
Vocabulary Size (Words)	32	128-512	255	64
Additional Characteristics	Speech output--32 words or phrases available	Board system for IBM PC's	Speech output available	Board system for IBM PC with 256K and disk drive
OPERATING CHARACTERISTICS				
Bandwidth (Hz)	na	na	150-4,000	na
Number of Filters	na	na	16	na
Bits Per Word	na	na	1000/0.5 sec	4000-14400/sec
Percent Recognition Accuracy	98	na	99+	na

VOICE SYNTHESIS SYSTEMS

MANUFACTURER AND MODEL	American Microsystems S3610/3620	Artic Technologies SynPhonix	Centigram Corp LISA	Centigram Corp TIM
GENERAL INFORMATION				
Purchase Price	$14-17 ea/1000 chips	$200	$2,450	$350
Typical Applications	Instrumentation, industrial controls, automobiles, games	Hobbyists, education	Computer response, training, handicapped, toys	Robots, education, alarms, toys
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	SP-0256 chip	IBM PC host	Z-80	Z-80, TI TMS5220 chip
System Memory	16K ROM	Uses host CPU	16K RAM, 32K PROM	ROM
System Software	Firmware	Disk	Proprietary	Proprietary
Communications Interface	IC Chip	Plug-in board	RS-232C	RS-232C, Multibus
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Phoneme synthesis, linear predictive coding, formant synthesis	Text-to-speech using phoneme synthesis	Parametric waveform coding	Linear predictive coding
Voices Supported	Male and female	na	na	na
Foreign Languages	None	None	na	na
Controls Supported	Volume, pitch, tone	na	na	na
		na	na	na
Vocabulary Size (Words)	Unlimited	na	Unlimited	Unlimited
Additional Characteristics	--	Includes SSi263 speech chip, amplifier, speaker	Available in Multibus version	Compatible with Multibus-based systems
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	na	na	na

MANUFACTURER AND MODEL	Centigram Corp VoiceWare	Digital Equip Co DECTalk	E-Systems CV-3670/A	General Instrument SP-0256 Orator
GENERAL INFORMATION				
Purchase Price	$25-29,500	na	na	Contact vendor
Typical Applications	Development, education handicapped	Office automation, computer response, handicapped	Militarized digital speech processors for U.S. D.O.D.	Automobiles, education, security systems, games, visually impaired
Number of Units Sold	na	na	1,400	na
PROCESSING FACILITIES				
System Processor	Z-80A	Any DEC computer	na	SP-0256 chip
System Memory	64K RAM, hard disk	Uses host CPU	na	16K ROM
System Software	Firmware	na	na	Firmware
Communications Interface	RS-232C, Multibus	RS-232C	na	IC chip
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Either LISA or TIM speech components	Text-to-speech with phoneme synthesis, coarticulation	Linear predictive coding	Phoneme synthesis, linear predictive coding, formant synthesis
Voices Supported	na	Male, female, child	No	na
Foreign Languages	na	Yes	No	None
Controls Supported	Rate, volume, pitch, tone	Speaking rate, pitch, breathiness, forte	na	Volume, pitch, tone
			na	
Vocabulary Size (Words)	Unlimited	20,000	na	Unlimited
Additional Characteristics	Development system	Pronounciation and intonation control	--	Can support up to 491K bits external ROM
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	96.7 MRT	na	na

VOICE SYNTHESIS SYSTEMS

MANUFACTURER AND MODEL	General Instrument VSM2128	Harris Semiconductor HC-55516/32/64	Interface Systems Freedom 1	Maryland Computer Information Thru Speech
GENERAL INFORMATION				
Purchase Price	$99	$8.25 ea/100	$500	$7,995-11,995
Typical Applications	Automobiles, education, security systems, games, visually impaired	Voice store-and-forward, phone messages, video games, voice message	Text-to-speech (from screen) for visually-impaired	Visually-impaired
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	GI PIC Series	IC chips	IBM PC host	HP125 using Z80A
System Memory	16K-128K ROM	10K-64K bits sec	64K on host	64K RAM
System Software	Firmware	Firmware	Disk	Firmware, disk
Communications Interface	Plug-in board	--	Serial port	RS-232C
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Phoneme synthesis, linear predictive coding, formant synthesis	Continuously variable-slope data modulation	Uses Votrax SC-01A speech synthesizer	Phoneme synthesis using Votrax VSM
Voices Supported	--	--	Male, female, robotic	--
Foreign Languages	None	Unlimited	None	None
Controls Supported	Volume, pitch, tone	Reflects true speaker characteristics	na	Volume, pitch, tone, rate
Vocabulary Size (Words)	Unlimited	Unlimited	Unlimited	Unlimited
Additional Characteristics	Uses GI SP-0256 chip	--	Requires 64K RAM, serial port, voice synthesizer chip	Talking Information Management System (TIM II) standard
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	10K-64K bits/sec	75-300 words/min	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	na	na	na

MANUFACTURER AND MODEL	Maryland Computer Speak Easy	Maryland Computer Total Talk	Micromint Inc Microvox	MIT MITalk-79
GENERAL INFORMATION				
Purchase Price	$3,000	$4,990	$349	na
Typical Applications	Manufacturing, alarms aircraft cockpits	Visually-impaired	Text-to-speech for visually-impaired	Research system at Massachusetts Institute of Technology
Number of Units Sold	na	Over 120	na	na
PROCESSING FACILITIES				
System Processor	na	HP2621A/dual Z-80s	6502	na
System Memory	na	ROM	8K ROM	na
System Software	na	Firmware	Firmware	na
Communications Interface	RS-232C, parallel	RS-232C	na	na
			na	na
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Phoneme synthesis using Votrax VSM	Phoneme synthesis using Votrax VSM	Uses Votrax SC-01A text-to-speech synthesizer	na
Voices Supported	--	--	Male tenor	na
Foreign Languages	None	None	No	na
Controls Supported	Volume, pitch, tone, rate	Volume, pitch, tone rate	Volume, pitch, rate	na
Vocabulary Size (Words)	Unlimited	Unlimited	na	na
Additional Characteristics	--	--	Can come in kit form	--
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	na	na	93.1 MRT

VOICE SYNTHESIS SYSTEMS

MANUFACTURER AND MODEL	Mimic Inc Mimic Speech Processor	National Semiconductor Digitalker I/II	Oracle Communications Oracle	Panasonic MN6401
GENERAL INFORMATION				
Purchase Price	$235	$20 ea/1000	na	na
Typical Applications	Alarms, games, aid for handicapped	Automotive, banking, weather, elevators, toys, alarms, games	na na na	Consumer products, games, education
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Apple II, TRS-80 host	MM54104/DT3101TD	na	NEC MN6401
System Memory	ROM	External ROM	na	32K ROM
System Software	Firmware	Firmware	na	Firmware
Communications Interface	Plug-in board	IC chips	na	IC chip
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Proprietary technique	Time-domain synthesis, speech compression	na na na	Partial auto correlation
Voices Supported	na	Male/female/child	na	na
Foreign Languages	na	na	na	na
Controls Supported	Volume, pitch, tone	Volume, pitch	na na	Speech rate
Vocabulary Size (Words)	Unlimited	Unlimited	na	Unlimited
Additional Characteristics	Can also be used for voice recognition	Model II compatible with I, uses less power	--	IC chip
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	na	na	na

MANUFACTURER AND MODEL	Sanyo LC8100	Street Electronics ECHO II/GP	Tecmar Inc Speech Master	Telesensory Systems Speech 1000/1010
GENERAL INFORMATION				
Purchase Price	na	$200	$495	$1,200-1,600
Typical Applications	na	Sound effects, robotics, education, handicapped	Text-to-speech for visually-impaired, education	Status, alarms, voice messaging, education, electronic mail
Number of Units Sold	na	Over 1000	na	na
PROCESSING FACILITIES				
System Processor	na	6502	IBM PC host	Intel 8085
System Memory	na	1K or 8K RAM	ROM	--
System Software	na	Disk	Disk	Proprietary
Communications Interface	na	Card slot in Apple RS-232C, parallel	Plug-in board	RS-232C, parallel, and multibus
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	na	Texas Instruments linear predictive coding, TMS5200 chip	Votrax SC-01 chip and National Semiconductor 'DigiTalker' LSI chip	Stored human speech, linear predictive coding,
Voices Supported	No	na	Male (danish accent)	Male or female
Foreign Languages	Yes, phrases	German	No	Unlimited
Controls Supported	na	Variable speech rate, pitch control	na na	Rate, volume, pitch
Vocabulary Size (Words)	96 words	Unlimited	143 words	Unlimited
Additional Characteristics	-	3 punctuation modes	Board system for IBM PC with 64K RAM and disk drive	Natural sounding, human quality
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	na	na	na	na

VOICE SYNTHESIS SYSTEMS

MANUFACTURER AND MODEL	Telesensory Systems Prose 2000/2020	Texas Instruments TMS5000 Series	Triformation Systems AVC-1	Triformation Systems Scan Speech Terminal
GENERAL INFORMATION				
Purchase Price	$3,500-4,800	Contact vendor	$1,895	$2,495 to $4,495
Typical Applications	Electronic mail, voice messaging, education	Most any application with a voice output requirement	Visually handicapped users	Visually handicapped users
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Intel 8086	TMS5100, 5110 & 5220	na	na
System Memory	na	32K or 128K ROM	ROM	2K RAM
System Software	Proprietary	--	Firmware	Firmware
Communications Interface	RS-232C and multibus	IC chips	RS-232C	RS-232C
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Text-to-speech, formant synthesis	Linear predictive coding	Text-to-speech, phoneme synthesis	na
Voices Supported	Male voice	Male or female	na	na
Foreign Languages	None	Unlimited	na	na
Controls Supported	Rate, volume, pitch	Rate, volume, pitch, tone	Volume, pitch, tone	Volume, tone
Vocabulary Size (Words)	Unlimited	200-3200 words	Unlimited	na
Additional Characteristics	Intonation rules, phoneme input	Natural sounding	Speaks entire words or spells out words	Speaks entire words or spells out words
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	na
Percent Recognition Accuracy	87.6 MRT	na	na	na

MANUFACTURER AND MODEL	Votrax Corp Speech Pac	Votrax Corp VSM/1	Votrax Corp Type-N-Talk	Vynet Corp V101/102 Series
GENERAL INFORMATION				
Purchase Price	Contact vendor	Contact vendor	$345	$695 to $1,995
Typical Applications	Alarm systems, elevators, education, toys	Industrial control, games, sound effects	Computer prompting, computer-aided instruction	Bank-by-phone, marketing, polling, database query
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Votrax SC-[illegible]	Votrax SC-01, M6800	Votrax SC-01	IBM PC host
System Memory	32K EPROM	1K RAM, 8K ROM	--	na
System Software	Firmware	Firmware	Firmware	Disk software
Communications Interface	Parallel	RS-232C, parallel	RS-232C	Plug-in board
				na
SPEECH CHARACTERISTICS				
Synthesis Technique(s)	Phoneme synthesis	Phoneme synthesis	Phoneme synthesis	Adaptive differential pulse code modulation (ADPCM)
Voices Supported	--	--	--	--
Foreign Languages	None	None	None	None
Controls Supported	Pitch, volume, tone	Pitch, volume, tone	Pitch, volume, tone	na
				na
Vocabulary Size (Words)	Unlimited	Unlimited	Unlimited	na
Additional Characteristics	Complete board level system	Complete microcomputer system	750-character buffer for typed characters	Board system for IBM PC, requires microphone and preamplifier
OPERATING CHARACTERISTICS				
Data Transmission Rate	na	na	na	na
Average Word Duration	na	na	na	na
Bits Per Word	na	na	na	4K bytes/sec
Percent Recognition Accuracy	na	na	66.2 MRT	na

VOICE RESPONSE SYSTEMS

MANUFACTURER AND MODEL	Amnet SolidVoice 4020	Cognitronics Model 681	Cognitronics Model 682	Cognitronics Model 683
GENERAL INFORMATION				
Purchase Price	$30,000	$3,000	$16,500	$6,800
Typical Applications	Data entry/retrieval, file update, interactive query/response, banking	Alarm announcements, security systems	Banking, order entry, Credit authorization	Banking, order entry, Credit authorization
Number of Units Sold	25	na	na	na
PROCESSING FACILITIES				
System Processor	na	na	na	na
System Software	Proprietary	Proprietary	Proprietary	Proprietary
Interaction	User prompts	--	Voice, tone	Voice, tone
No of Communication Lines	Up to 48	3 channels	Up to 48	Up to 12
Remote Terminals Supported	Touch-tone input, alphanumeric terms	Output only	Touch-tone input	Touch-tone input
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases	Words, phrases	Words, phrases
Size	1,000 words/message	64 words/message	254 words/message	254 words/message
Generation Technique(s)	Continuous variable-slope delta modulation (CVSD)	Delta modulation	Delta modulation (CVSD)	Delta modulation (CVSD)
System Configuration	Subsystem on-line to host	Front-end	Front-end	Front-end
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	2,880	--	1,440 (24 lines)	360 (6 lines)
3-minute calls	960	--	480 (24 lines)	120 (6 lines)

MANUFACTURER AND MODEL	Cognitronics VRCS/SL	Computer Consoles Audio Response System	Computer Curriculum Digital Speech System	Computer Curriculum Telephone Access Unit
GENERAL INFORMATION				
Purchase Price	$65,000	Contact vendor	$9,000 to $10,000	$5,000 to $7,500
Typical Applications	Banking, order entry credit authorization	Telephone directory assistance	Computer-aided instruction	Computer-aided instruction
Number of Units Sold	na	na	40	na
PROCESSING FACILITIES				
System Processor	na	na	na	na
System Software	Proprietary	Proprietary	Proprietary	Proprietary
Interaction	Voice, tone	Operator-controlled	Prompts via host	Prompts via host
No of Communication Lines	Up to 12	na	1-5	1-5
Remote Terminals Supported	Touch-tone input	Directory assistance consoles	Alphanumeric terminals	Alphanumeric terminals, touch-tone input
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases	Words, phrases	Words, phrases
Size	254 words/message	na	Unlimited	Unlimited
Generation Technique(s)	Delta modulation	Digitized speech	Linear predictive coding	Linear predictive coding
System Configuration	Freestanding system	Component within telephone directory assistance system	Freestanding; 16-terminal cluster subsystem	Freestanding; 6- to 12-terminal cluster subsystem
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	360 (6 lines)	na	360-720	360-720
3-minute calls	120 (6 lines)	na	120-240	120-240

VOICE RESPONSE SYSTEMS

MANUFACTURER AND MODEL	Digital Pathways SLC-II	Engineered Systems TCS/500 Series	IBM 7770	Infolink 1500/1550 CVR
GENERAL INFORMATION				
Purchase Price	$2,800	$20,000 to $243,000	$58,760 and up	$15,000 to $60,000
Typical Applications	Computer monitoring, data entry/retrieval, security	Transaction processing, banking, data entry, order entry	--	Order entry, telephone bill payment, banking, inquiry status
Number of Units Sold	200	na	na	140/60
PROCESSING FACILITIES				
System Processor	Proprietary 6502-based	TI 990 Series	IBM hard-wired	DEC PDP-11/23
System Software	Basic, firmware	Proprietary	Assembly	Custom-designed
Interaction	Prompting	Voice, tone, ASCII code	Prompts	Prompts
No of Communication Lines	4 RS-232C, 1 telephone	1 to 72	4 to 48	8/32
Remote Terminals Supported	Touch-tone input, ASCII modems	Touch-tone input, transaction phone, CRTs, ATMs, POS	Touch-tone input various IBM terminals	Touch-tone input, transaction phone
SPEECH CHARACTERISTICS				
Type	Words	Words, phrases, sent.	Words	Words, phrases
Size	Unlimited	640 words	128-word vocabulary	Up to 392 words/msg
Generation Technique(s)	Linear predictive coding	Linear predictive coding	Amplify recorded signal	Digitized voice
System Configuration	Freestanding; RS-232C connection	Down-line loaded remote system; front-end for mini/main	Subsystem under host control	Subsystem under host control
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	60	60-4,320	60-2,100	na
3-minute calls	20	20-1,440	20-700	na

MANUFACTURER AND MODEL	Interface Technology TOES	Interstate Electronics VTM 150	Master Specialties Models 1650/1700/1750	Natural Microsystems Watson
GENERAL INFORMATION				
Purchase Price	$60,000 and up	Contact vendor	$2,000 and up	$998
Typical Applications	Order entry, sales reports, payroll reporting, credit autho.	Data entry	Aircraft warning, telecommunications, paging, credit verification	Voice messaging, voice storage, hobbyists
Number of Units Sold	14	na	na	na
PROCESSING FACILITIES				
System Processor	IBM Series/1	Proprietary	Z-80, others	IBM PC host
System Software	Total entry system	Proprietary in EPROM	Prop. in PROM, ROM	
Interaction	Prompting - 3 levels	12 commands	--	User ID code
No of Communication Lines	1-32	Up to 28	1	1
Remote Terminals Supported	Touch-tone input CRTs, POS terminals	Alphanumeric terminals	--	Touch tone input
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases	Words, phrases	Words, phrases, sentences
Size	Unlimited	1,500-word vocabulary	Up to 256 sec./msg	Depends on disk size
Generation Technique(s)	Continuously variable-slope delta modulation	Phoneme synthesis	Digitized voice, delta modulation	Delta modulation
System Configuration	Freestanding; front-end	Component within computer system	Subsystem under host control	Requires 192K RAM on host and two disk drives
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	4,320	na	na	na
3-minute calls	1,440	na	na	na

VOICE RESPONSE SYSTEMS

MANUFACTURER AND MODEL	Perception Technology BT-II	Perception Technology VOCOM I	Perception Technology VOCOM II	Periphonics T-Comm 80
GENERAL INFORMATION				
Purchase Price	$15,000 to $50,000	$45,000 and up	$60,000 and up	$60,000 and up
Typical Applications	Cash balance reporting, payroll & order entry, credit authorization	Cash balance reporting, payroll & order entry, credit authorization	Cash balance reporting, payroll & order entry, credit authorization	Banking, data entry, security
Number of Units Sold	Over 250	123	29	Over 250
PROCESSING FACILITIES				
System Processor	DEC LSI-11/02	DEC LSI-11	DEC PDP-11, LSI-11/02	Proprietary
System Software	Proprietary	Proprietary	na	Proprietary
Interaction	Prompt, repeat input, invalid ent., host	Prompt, repeat input, invalid ent., host	Prompt, repeat invalid ent., host	Voice
No of Communication Lines	4-32, asynchronous	4-32, asynch/bisync	4-256	3-104
Remote Terminals Supported	Touch-tone input, POS terminals	Touch-tone input, POS terminals	Touch-tone input, POS terminals	Touch-tone input, POS terms, CRTs badge readers
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases	Words, phrases	Words, phrases
Size	Up to 992 seconds	Up to 992 seconds	Up to 992 seconds	1,500 words/message
Generation Technique(s)	Adaptive delta modulation	Adaptive delta modulation	Adaptive delta modulation	na
System Configuration	Freestanding front-end processor - emulates terminal interface	Freestanding front-end processor - emulates terminal or controller	Standalone data base system or RJE to mainframe (2780/3780)	Freestanding system, subsystem, remote system
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	240-1,920	240-1,920	192-1,536	180-6,240
3-minute calls	80-640	80-640	64-512	60-2,080

MANUFACTURER AND MODEL	Telesensory Speech T/VIS	Votan Corp V4000/5000	Votrax Corp LVM-80	Votrax Corp VS-6
GENERAL INFORMATION				
Purchase Price	$11,200 to $72,000	$3,200 to $5,000	$8,750 and up	$1,100 to $4,050
Typical Applications	Electronic mail via telephone, order entry, information inquiry	Computer-aided instruction, data entry/retrieval	Business communication, data entry	Data entry
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Intel 8086	2901 bit-slice CPUs	Proprietary, VODIS-1	Proprietary
System Software	Proprietary	Proprietary	Proprietary	Proprietary
Interaction	Prompts via host	CRT	Voice prompts	Data terminal inputs
No of Communication Lines	1-12	na	4-64	na
Remote Terminals Supported	Touch-tone input	na	Touch-tone input, transaction telephone	Touch-tone input Transcom 160
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases	Words, phrases	Words, phrases
Size	Unlimited	Depends on bit rate	Up to 512 seconds	Unlimited
Generation Technique(s)	Linear predictive coding or text-to-speech	Bit rate spectral coding	Digitized voice	Phoneme synthesis
System Configuration	On-line to host CPU	Board-level; freestanding system	Freestanding system	Remote system connected to host CPU
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	240-720	na	240-3,480	60
3-minute calls	80-240	na	80-1,280	20

VOICE RESPONSE SYSTEMS

MANUFACTURER AND MODEL	Votrax Corp ML-1	Vynet Corp V101/102 Series		
GENERAL INFORMATION				
Purchase Price	$2,000 to $8,000	$695 to $1,995		
Typical Applications	Data entry tion, data entry	Order entry, stock control, remote process, status enquire		
Number of Units Sold	na	na		
PROCESSING FACILITIES				
System Processor	Proprietary	Uses host CPU		
System Software	Proprietary	Proprietary		
Interaction	Data terminal inputs	Prompts via host		
No of Communication Lines	na	1-2		
Remote Terminals Supported	Touch-tone input, Transcom 160	Touch-tone input		
SPEECH CHARACTERISTICS				
Type	Words, phrases	Words, phrases		
Size	Unlimited	Unlimited		
Generation Technique(s)	Phoneme synthesis	Adaptive differential pulse code modulation (ADPCM)		
System Configuration	Remote system connected to host CPU	Subsystem for Apple II or IBM PC		
OPERATING CHARACTERISTICS				
Message Capacity, calls per hr				
1-minute calls	60	60		
3-minute calls	20	20		

VOICE MANAGEMENT SYSTEMS

MANUFACTURER AND MODEL	AT & T Voice Store/Forward	BBL Industries Voice Retrieval System	Centigram Corp Voice Memo	Commterm Inc EVX 500
GENERAL INFORMATION				
Purchase Price	Contact vendor	$60,000 to $250,000	$21,600 to $55,300	$15,000 to $92,000
Basic Configuration	Freestanding or integrated with Dimension PBX FP11 and 8.2	Freestanding; connects to PABX	Freestanding; connects to PABX	Freestanding, message service, or integrated with PABX
Number of Units Sold	na	130	50	na
PROCESSING FACILITIES				
System Processor	IBM Series 1 4956	Z-80	Intel 8088	Z-80
System Memory	512KB	na	256KB	32K ROM, 144K RAM
Hard Disk Subsystem	1100MB	33MB to 600MB	140MB	None
System Software	ADS Appli. Software	Proprietary	Proprietary	Prop., Z-80 Assembler
Other Peripherals Supported	Printer, display stn	Printer, CRT, computer interface	Printer, CRTs, DID interface	CRT, printer, PABX serial port
INPUT/OUTPUT				
Communication Lines	16 ports	2/64	1/14 ports	2/8
Input Devices Supported	Touch-tone phone	TT or rotary phone	Touch-tone phone	TT or rotary phone
Message Waiting Notification	Light/ringback	No	Special ringback	Dialout
System Interaction	Voice prompts	Custom voice prompts	Voice prompts, signals	Voice prompts, tones
SPEECH CHARACTERISTICS				
Type Supported	Sentences	Words, sentences, etc.	--	Continuous speech
Generation Technique(s)	Digitized voice	CVSD digitized voice	--	CVSD, 20K bps
OPERATING CHARACTERISTICS				
Number of Users	2,000 maximum	64,000 maximum	1/1000	50/500
Max Number of Messages	6,984	63 per subscriber	45 mailbox	3000
Message Length	52 sec to 56 min	5 sec to 10 minutes	Variable	Variable, 7 min max
Hours of Message Stored	100 hours	40 hours	14 hours	17 hours

MANUFACTURER AND MODEL	CommTerm Inc EVX 1000	CommTerm Inc EVX 2000	CommTerm Inc EVX 9000	CommTerm Inc EVX TRA
GENERAL INFORMATION				
Purchase Price	$33,000 to $175,000	$95,000 to $200,000	$50,000 to $500,000	$58,000 to $218,000
Basic Configuration	Freestanding, message service or integrated with PABX	Freestanding, message service or integrated with PABX	Freestanding, message service or integrated with PABX	Freestanding, message service, interface to to paging, mobile phone
Number of Units Sold	40	na	na	32
PROCESSING FACILITIES				
System Processor	Z-80	Z-80	M68000	Z-80
System Memory	32K ROM, 144K RAM	32K ROM, 144K RAM	32K RAM	64K RAM
Hard Disk Subsystem	None (floppy)	None (floppy)	5-20MB	5MB
System Softwarea	Prop., Z-80 Assembler	Prop., Z-80 Assembler	Prop., C Langauge	Prop., Z-80 Assembler
Other Peripherals Supported	CRT, printer, PABX serial port	CRT, printer, PABX serial port	System console device, PABX serial port	System console device
INPUT/OUTPUT				
Communication Lines	2/32	2/32	2/64	2/32
Input Devices Supported	TT and rotary phone	TT and rotary phone	TT and rotary phone	TT and rotary phone
Message Waiting Notification	Depends on PABX	Depends on PABX	Depends on PABX	Outdial
System Interaction	Voice prompts, tones	Voice prompts, tones	Voice prompts, tones	Voice prompts, tones
SPEECH CHARACTERISTICS				
Type Supported	Continuous speech	Continuous speech	Continuous speech	Continuous speech
Generation Technique(s)	CVSD, 20K bps	CVSD, 20K bps	CVSD, 20K bps pause detect	CVSD, 20 K bps
OPERATING CHARACTERISTICS				
Number of Users	50/1000	50/2000	100/1000	50/5000
Max Number of Messages	5000	6000	na	16,000
Message Length	Variable, 7 min max	Variable, 7 min max	Variable, 20 min max	160 sec
Hours of Message Stored	60 hours	120 hours	180 hours	60 hours

VOICE MANAGEMENT SYSTEMS

MANUFACTURER AND MODEL	Ericsson Inc VOIS	GTE Business Systems VNX-III	IBM Series I Audio Distribut'n System	Northern Telecom SL-1
GENERAL INFORMATION				
Purchase Price	na	na	$44,336 and up	na
Basic Configuration	Integrated with ASB 900	Integrated with PABX (GTD-4600)	Freestanding, connects to PABX	Integrated with SL-1
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Z-80	Intel 8086	IBM Series 1	Z-80
System Memory	32K ROM, 144K RAM	256K-512K RAM	na	32K ROM, 144K RAM
Hard Disk Subsystem	None (floppy)	330-1896MB	na	None (floppy)
System Software	Proprietary	Proprietary	IBM licensed program	Prop., Z-80 Assembler
Other Peripherals Supported	CRT, printer, PABX serial port	Printer, CRT, RS-232C	na	CRT, printer, PABX serial port
INPUT/OUTPUT				
Communication Lines	2/32	8/64	2/10	2/32
Input Devices Supported	TT and rotary phone	Touch tone phone	Touch tone phone	TT and rotary phone
Message Waiting Notification	Depends on PABX	Page, lamp, tone, msg.	na	Depends on PABX
System Interaction	Voice prompts, tones	Voice prompts, tones	Voice	Voice prompts, tones
SPEECH CHARACTERISTICS				
Type Supported	Continuous speech	Sentences	Sentences	Continuous speech
Generation Technique(s)	CVSD, 20K bps	Digitized voice	Digitized voice	CVSD, 20K bps
OPERATING CHARACTERISTICS				
Number of Users	50/1000	500 to 8000	na	50/1000
Max Number of Messages	5000	10200 on-min messages	1630 per day (1 min)	5000
Message Length	Variable up to 7 min	Variable, 5 min max	Variable	Variable, 7 min max
Hours of Message Stored	60 hours	170 hours	na	60 hours

MANUFACTURER AND MODEL	Octel Communications Aspen	Optimum Communica'n Opcom VX	Rolm Inc PhoneMail System	Solid State Systems Electronic Message Bank
GENERAL INFORMATION				
Purchase Price	$55,000	$20,000 to $40,000	$70,000 to $185,000	$100,000 to $500,000
Basic Configuration	Freestanding, integrates with PABX and Centrex	Freestanding, connects to PABX	Integrated with Rolm CBX	Freestanding, integrates with PABX
Number of Units Sold	na	na	na	na
PROCESSING FACILITIES				
System Processor	Proprietary	Microprocessor	Intel 8086	Proprietary
System Memory	1MB to 3MB	400KB	1MB RAM, ROM	na
Hard Disk Subsystem	80MB to 480MB	4-20 hours	7-154MB	na
System Software	Proprietary	Proprietary	Proprietary	na
Other Peripherals Supported	Terminal or printer	Printer, CRT, modem	1200-baud printer/term	na
INPUT/OUTPUT				
Communication Lines	4/24	4/16	8/16 channels	64 maximum
Input Devices Supported	TT or rotary phone	Touch tone phone	Touch tone phone	TT or rotary phone
Message Waiting Notification	Notification call	Ringing delivery	Lamp, LED, tones	na
System Interaction	Voice prompts	Voice prompts	Voice prompts	Voice prompts
SPEECH CHARACTERISTICS				
Type Supported	Sentences	Sentence	Sentences	Sentences
Generation Technique(s)	Digitized voice	Digitized voice	CVSD	Digitized voice variable rates
OPERATING CHARACTERISTICS				
Number of Users	5/3000	Up to 1,000	50/1000	3000 per day
Max Number of Messages	465,000	Up to 65,000	No system limit	na
Message Length	Variable by class	Unlimited	Varies per user	Variable
Hours of Message Stored	43 hours	4 to 20 hours	51 hours	na

VOICE MANAGEMENT SYSTEMS

MANUFACTURER AND MODEL	Sperry Voice Info Processing	VMX Inc Voice Message Exchange	Voicemail Intern'al VIX/64	Voicetek Corp VoiceStar
GENERAL INFORMATION				
Purchase Price	$126,000 to $210,000	$92,500 to $525,000	$441,180	na
Basic Configuration	Freestanding, integrates to PABX or Sperrylink	Freestanding, connects to PABX	Freestanding	Freestanding, connects to PABX
Number of Units Sold	na	59	na	na
PROCESSING FACILITIES				
System Processor	Proprietary	Intel 8086	PDP-11/24	DG Micro Nova
System Memory	256K to 512K RAM	256K to 512K RAM	2MB RAM	na
Hard Disk Subsystem	na	330MB-1896MB	3192MB	na
System Software	Proprietary	Proprietary	Proprietary VMOS	Proprietary
Other Peripherals Supported	na	Printer, RS-232C	na	na
INPUT/OUTPUT				
Communication Lines	2/12	8/64	64 ports	2/32
Input Devices Supported	Touch tone phone	Touch tone phone	TT or rotary phone	TT or rotary phone
Message Waiting Notification	na	Ringback, lamp, pager	Callout	na
System Interaction	Voice prompts	Voice prompts	Voice prompts, DTMF	Voice prompts
SPEECH CHARACTERISTICS				
Type Supported	Sentences	Sentences	Continuous speech	Sentences
Generation Technique(s)	Digitized voice	Digitized voice	Digitized voice, 32 kb	Continuous variable-slope delta modulation (CVSD)
OPERATING CHARACTERISTICS				
Number of Users	na	500 to 8000	13,000	na
Max Number of Messages	na	10,200 one-minute	6000 15-sec messages	12,500
Message Length	na	5 minutes maximum	15 sec to 5 min	na
Hours of Message Stored	32 hours	na	25 hours	na

MANUFACTURER AND MODEL	Votan Corp Voice Mgmt System	Wang Laboratories DVX		
GENERAL INFORMATION				
Purchase Price	$35,000	$90,250 to $186,250		
Basic Configuration	Freestanding	Freestanding, connects to PABX		
Number of Units Sold	na	97		
PROCESSING FACILITIES				
System Processor	8086 family	Wang DVX		
System Memory	1MB RAM	na		
Hard Disk Subsystem	40MB-70MB	na		
System Software	RMX-86	Proprietary		
Other Peripherals Supported	Printer, CRT terminal	na		
INPUT/OUTPUT				
Communication Lines	1/8	4/16		
Input Devices Supported	TT and rotary phone	Touch tone phone		
Message Waiting Notification	na	na		
System Interaction	Voice prompts	Voice prompts		
SPEECH CHARACTERISTICS				
Type Supported	Continuous speech	Sentences		
Generation Technique(s)	Digitized voice	Digitized voice		
OPERATING CHARACTERISTICS				
Number of Users	8 simultaneous	na		
Max Number of Messages	Unlimited	360 to 3140 per day		
Message Length	Variable	Variable, 6 min max		
Hours of Message Stored	Depends on disk cap.	na		

The accompanying tables summarize the functional requirements and physical and operational solutions for various modes of transportation for the elderly and disabled. The tables have been divided into seven distinct areas: special E&H transit, urban bus transit, urban rail transit, inter-urban bus, inter-urban rail, and a typical transportation terminal. The type of system is indicated at the top of each page. The tables have been basically derived from the report "Transportation Problems of the Transportation Handicapped" by the United States Urban Mass Transportation Administration, August 1976. The table entries and their significance are explained in the following paragraphs.

Functional Requirements. The tables start off by listing all of the functional requirements that must be undertaken by individuals in order to use a particular transportation mode, such as obtaining information on the service provided and accessing the vehicles themselves.

Physical Solutions. Various physical or technological solutions that would either assist the elderly and disabled in meeting the functional requirement or remove the functional requirement are then presented. The areas where the solution might take advantage of voice technology are shaded.

Operational Solutions. Where an operational solution is feasible, these are indicated.

SPECIAL E&H TRANSIT

Functional Requirement	Physical Solution	Operational Solution
Register for service	Fill out forms in person or by phone to confirm eligibility for service and to provide information for operator to provide personal service	Vehicle driver confirmation of eligibility
Reserve/book a trip	Phone operator to book a trip; possibly use voice response units to log the information	
Confirm status/times	Automatic callback controllers, audio response detectors and voice synthesis message units	
Confirm vehicle arrival at door	Larger bus markings, audio cue to indicate arrival, automatic callback controllers with voice synthesis message units to announce imminent arrival	
Reach bus entrance	Curb cuts, ramps, elevators	Driver assistance
Fare transaction manipulation	Use of large tickets; operator billing of trips on monthly basis	Free fare policy
Take position in vehicle, secure safety devices	Low-effort mechanical latches and straps	Driver assistance
Exit vehicle	Low-effort mechanical latches and devices	Driver assistance

URBAN BUS TRANSIT

Functional Requirements	Physical Solution	Operational Solution
Go 2 blocks		Demand-responsive, non-fixed-route service; or fixed route with deviations
Negotiate curbs	Curb cuts	
Cross street	Special crossing lights operated by E&H card	
Identify bus stop	Standardized high-visibility day-night bus stop markings; "referral signs"; Broader use of audio, visual media for route, time information	
Identify arrival time	Better, broader use of audio, visual cues for time information; Better telephone information service for arrival times	
Wait standing	Shelters with seating, wheelchair space, rain protection, heat, splash protection, security	
Identify correct bus	Larger bus markings; PA announcement; Talking bus stop	
Reach bus entrance	Sound cue to indicate door position	
Step up to bus bottom step	Eliminate need (high platform operation); 3-way steps/lift/platform that lowers to curb or street level; kneeling bus	

URBAN BUS TRANSIT (cont'd)

Functional Requirement	Physical Solution	Operational Solution
Pass through bus door	Clear door openings greater than 32"	
Climb bus steps	Eliminate need (high platform operation); 3-way steps/lift/platform device in bus; lower bus floor, have fewer steps	
Grasp handrail	Properly placed grips with non-slip surfaces	
Fare transaction manipulation	Out-of-vehicle fare transaction; Honour system ticket cancellation by passenger	Free fare for identified E&H
Identify vacant seat or standing space	Make special seating area (in consistent location for E&H only)	Enforced E&H priority selec- of any seat
Pass along aisle in moving vehicle	Eliminate need (reserve E&H seating near entrance); Wider aisle; Provide adequate grips and stanchions	Institute "No-stand" policy
Grasp vertical stanchion; Grasp overhead grip; Grasp seatback grip	Provide sufficient of each to give rider alternative; Make all grips large enough to encourage sharing	
Sit in lateral seat; Sit in forward-facing aisle seat	Provide each as alternative to other; Provide grips and arms to assist sitting, rising in both	
Pass person in aisle seat	Enlarge foot room; Provide grips	Provide reserved seating
Sit in window seat	Eliminate need (designate reserved seating); use bench seats to encourage other riders to move over; Provide grips, arms	Provide reserved seating for E&H
Ride standing	Eliminate need (designate reserved seating for E&H); Provide adequate stanchions and non-slip grips	Enforced E&H priority for any seat; Institute "No-Stand" policy
Adjust position to let others by	Provide grips and seat arms to assist, steady move- ment	
Identify desired stop	Adequate destination markings and announcements	
Reach buzzer cord	Eliminate need by providing reachable pressure- operated signal	
Rise from seat	Provide seat arms, grips	
Pass through aisle in mov- ing vehicle	Eliminate need (reserve E&H seating near entrance); Wider aisle significantly; Provide adequate grips and stanchions	"No-Stand" policy (passen- gers move only when vehicle stationary)
Move to door area	Designate E&H seating near doors; Provide grips and stanchions	
Ride standing during deceleration	Adequate grips and stanchions	"No-Stand" policy
Go down bus steps	Eliminate need (high platform operation); Provide 3-way steps/lift/platform device in bus; Lower bus floor, have fewer steps	
Pass through door	Clear door openings greater than 32"	
Step down to ground or across to curb from bottom bus step	Eliminate need (high platform operation); Provide lift that reaches to curb or street level; Have kneeling bus	

URBAN RAIL TRANSIT

Functional Requirement	Physical Solution	Operational Solution
Go 2 blocks		Accessible door-to-door special service feeders
Negotiate curbs	Curb cuts	
Cross streets	Multiple station entrances on all corners of intersection or on both sides of street	
Identify station	Standardized terminal symbols; "referral" signs	
Approach fare barrier	Channelization; better use of audio, visual, textural media for in-terminal path and time information	
Take position in file	Channelization; special fare gate for E&H	
Manipulate currency or "proof-of-payment"	Improved detail design of fare barrier	Free fare for E&H
Activate barrier gate	Design of fare barrier: "always open"; lower-effort turnstyle	Provide alternative attended gate for E&H
Pass through fare barrier	Minimum clear channel width of 32"	
Identify direction to boarding area	Channelization; broader use of audio, visual, textural media to indicate location	
Go 20 ft. to 100 yards	SEE "Representative Terminal"	
Avoid hazard (open track pit)	Platform edge barrier; platform edge delineation	
Identify correct train	Better route identification on vehicles	
Observe door position	Improved train braking control to "spot" door within narrow limits; sound, light are at vicinity of open door	
Go to vicinity of door	Priority reserved waiting area for identified E&H persons; channelization	
Move in crowd	Priority reserved waiting area	
Cross car-platform floor discontinuity	Automatic adjustment of car suspension; gap-bridging devices	
Pass through door	Clear door opening greater than 32"	
Identify vacant seat	Eliminate need (designate E&H seats near door)	
Pass along aisle	Eliminate need (provide reserved seats, wheelchair space near door); widen aisle; provide adequate stanchions and grips	
Grasp vertical stanchion; Grasp overhead grip; Grasp seatback grip	Provide sufficient of each to give rider alternative; make all grips large enough to encourage sharing	
Sit in lateral seat; Sit in forward-facing aisle seat	Provide each as alternative to other; provide grips and arms to assist sitting, rising in both	
Pass person in aisle seat	Enlarge footroom; provide grips	Provide reserved seating for E&H
Sit in window seat	Eliminate need (designate reserved seating); use bench seats to encourage other riders to move over; provide grips, arms	Provide reserved seating for E&H

URBAN RAIL TRANSIT (cont'd)

Functional Requirement	Physical Solution	Operational Solution
Ride standing	Eliminate need (designate reserved seating for E&H) provide adequate stanchions and non-slip grips	Enforced E&H priority for any seat; institute "No-Standing" policy
Adjust position to let others by	Provide grips and seat arms to assist, steady movement	
Identify desired stop	Adequate destination markings and announcements	
Reach buzzer cord	Eliminate need by providing reachable pressure-operated signal	
Rise from seat	Provide seat arms, grips	
Pass through aisle in moving vehicle	Eliminate need (reserve E&H seating near entrance); wider aisle significantly; provide adequate grips and stanchions	"No-Stand" policy (passengers move only when vehicle is stationary)
Move to door area	Designate E&H seating near doors; provide grips and stanchions	
Ride standing during deceleration	Adequate grips and stanchions	"No-Stand" policy
Rise from seat	Provide grips and arms	
Move to door area	Eliminate need (priority seating near door)	
Coordinate with other passengers	Channelization debarking passengers in vehicle	
Pass through door	Clear door opening greater than 32:	
Cross car-platform floor discontinuity	Automatic adjustment of car suspension; gap-bridging devices	
Move safe distance from departing train, track pit	Delineate platform edge (lights, textures)	

INTER-URBAN BUS

Functional Requirement	Physical Solution	Operational Solution
Go 2 blocks		Provide accessible feeder modes
Negotiate curb	Curb cuts	
Cross street	Operate crossing lights with special E&H card	
Identify terminal	Standardized terminal symbols; "referall" signs	
Identify boarding area	Broader use of audio, visual media for route, terminal path, and time information	
Wait 60 minutes	Provide heated waiting room with accessible rest rooms	
Identify boarding time	Broader use of audio, visual media for time information	
Move to boarding location	Channelization; prioritized waiting area for E&H	
Step up to first step of bus	Eliminate need (high platform operation); match station curb height to bus step	
Pass through bus door	Clear door opening greater than 32"	
Climb bus steps	Eliminate need (high-platform operation); provide 3-way steps/lift platform device in bus; extend platform into bus from raised boarding area; design "jetway" for bus	
Fare transaction	Ticket sales in terminal; "honour system"	Free fare for E&H
Pass along aisle in vehicle	Eliminate need (provide marked, reserved seats for E&H near door)	
Identify vacant seat	Eliminate need (provide marked, reserved seats for E&H)	
Sit in forward-facing seat aisle seat	Provide grips and seat arms to assist	
Move past person in aisle	Enlarge foot room; provide grips	
Sit in window seat	Provide grips, arms	
Adjust position; control movement	Provide grips, arms	
Operate personal comfort devices	Place controls within waist to shoulder height of seated person	
Use rest room	Provide accessible rest room near entrance and/or reserved seats	
Identify stop	Adequate destination markings and announcements	
Rise from seat	Provide grips, arms	
Enter, move along aisle	Eliminate need (reserved seating near door); wider aisles; adequate grips and stanchions	
Go down bus steps	Eliminate need (high platform operation); provide 3-way steps/lift/platform device in bus; lower bus floor, have fewer steps	
Pass through door	Clear door openings greater than 32"	
Step down to ground or across to curb from bottom bus step	Eliminate need (high platform operation); provide list that reaches to curb or street level; have kneeling bus	

INTER-URBAN RAIL

Functional Requirement	Physical Solution	Operational Solution
Identify boarding point	Broader use of audio, visual media for in-terminal directional information	
Move to boarding position	Channelization; moving sidewalks; improved accessibility within terminals	
Wait 60 minutes	Heated waiting room with seats, accessible rest rooms	
Identify correct train and boarding point	Better route identification on vehicles; announcements	
Identify correct time to board	Broader use of audio, visual media for schedule, time information	
Cross a level surface	Moving sidewalk; in-terminal vehicle	
Step up to bottom train step	Eliminate need (high platform operation)	
Cross discontinuity between platform and car floor	Adjust car suspension automatically; bridge gap with device	
Pass through exterior door	Clear door opening greater than 32"	
Grasp, operate compartment door handle	Electric push-button latch; low-effort mechanical latch	Have all doors open in station
Open compartment door	Automatic and/or powered door	
Pass through compartment door	Clear door opening greater than 32"	
Identify vacant seat	Eliminate nee (reserve E&H seats near door)	
Move along aisle	Widen aisles	
Sit in aisle seat	Provide arms, grips	
Move past person in aisle seat	Enlarge footroom; provide grips	
Sit in window seat	Provide grips, arms	
Adjust position; control movement	Provide grips, arms	
Operate personal comfort devices	Place controls within waist to shoulder height of seated person	
Use rest room	Provide accessible rest room near entrance and/or reserved seats	
Fare transaction with conductor	Advance out-of-vehicle confirmation of ticket; provide stick-on confirmation	Eliminate fare for E&H
Identify stop	Adequate destination markings and announcements	
Rise from seat	Grips, arms	
Move past person in aisle	Increased foot space	
Enter, move along aisle	Eliminate need (reserve E&H seats near door); provide grips, stanchions	
Grasp, operate door handle	Electric push-button/low-effort mechanical latch	
Open compartment door	Powered door	

INTER-URBAN RAIL (cont'd)

Functional Requirement	Physical Solution	Operational Solution
Pass through compartment	Clear door opening greater than 32"	
Descend steps	Eliminate need (high platform operation); provide 3-way lift steps/platform in train car	
Pass through exterior door	Clear door openings greater than 32"	
Step down to track level	Eliminate need (high platform operation); provide 3-way lift steps/platform in car that lowers to track level	
Cross discontinuity between car floor and platform	Adjust train suspension; bridge gap with in-car or platform device	

INTER-URBAN AIR

Functional Requirement	Physical Solution	Operational Solution
Go 2 blocks		Accessible direct intermodal feeders
Negotiate curbs	Curb cuts	
Cross street	Signals activated by E&H card	
Identify terminal	Standardized terminal symbols	
Pass through metal detector	Widen passage to 32" (wheelchair will set off detector)	
Present ticket	Stick-on ticket envelopes	
Discuss seat preference		
Move into waiting area	All openings wider than 32"	
Wait seated or standing	Provide seats and wheelchair space; priority boarding service and waiting area; accessible rest rooms	
Identify boarding time	Better, broader use of audio, visual cues for time information	
Identify boarding path	Better, broader use of audio, visual cues for directional information	
Negotiate level change	Provide alternatives: steps, ramps, elevator	
Pass through jetway	Handrails in jetway; minimize floor slant	
Go 60 feet in jetway		
Negotiate 1 inch step	Ramp aircraft floor to sill	Adjust jetway carefully
Go 40 ft on paved surface	Eliminate need; provide jetway	
Identify correct aircraft	Broader use of audio, visual media to identify boarding path	

INTER-URBAN AIR (cont'd)

Functional Requirement	Physical Solution	Operational Solution
Climb boarding stairs	Eliminate need (provide jetway or alternate lifting device)	
Turn sharply inside aircraft	Place partitions to permit wheelchair maneuver	
Hand over boarding pass	Adhesive boarding pass adheres	
Enter, move along aisle	Eliminate need (priority E&H seating near door); narrow wheelchair; widen aisle; provide grips	
Identify vacant seat		
Sit in aisle seat	Grips, arms	
Move past person in aisle seat	Eliminate need (reserved E&H aisle seats); widen footspace; grips	
Sit in middle or window seat	Eliminate need (reserve E&H aisle seats); provide grips, seat arms	
Adjust position	Grips, arms	
Fasten seatbelt	Belt holder to keep belt ends accessible	
Perceive, understand safety lecture		
Tolerate acceleration	Keep acceleration within tolerable limits	
Operate personal environmental equipment	Place controls between waist and shoulder height of seated person	
Tolerate air pressure changes		Operate cabin pressurization at higher level
Reach, use rest room	Put rest room near reserved E&H seating; follow accessible guidelines for rest rooms	
Perceive warnings (e.g. fasten seatbelts)	Broader use of audio, visual media for these warnings	
Unfasten seatbelt	Push button release, auto retraction	
Rise from seat	Provide grips, arms	
Enter, move through aisle	Eliminate need (reserved seats near door); wider aisle, grips	
Turn sharply	Partitions arranged to permit wheelchair movement	
Pass through aircraft door	Clear door opening greater than 32"	
Cross door sill	Ramp aircraft floor to sill	
Go 60 ft. in jetway	Minimize jetway slant; provide handrails	
Cross 1" step in jetway	Probably not difficult for most E&H	
Descend boarding stairs	Eliminate need (jetway); platform lift device	
Identify correct gate	Broader use of audio, visual media to indicate pathway	
Go across paved surface	Eliminate need (jetway)	Wheelchair attendant
Pass through doorway	Clear door opening greater than 32"	

TYPICAL TERMINAL

Functional Requirement	Physical Solution	Operational Solution
Identify, locate terminal entrance	Standardized terminal symbols; "referral signs"	
Operate terminal door	Automatic door	
Pass through entrance door	Clear door opening greater than 32"	
Perceive, understand informational media	Broader, better use of audio and visual media for route, direction, and time information	
Fare transaction	Detail design to reduce intricate money or ID manipulation	Free fare for handicapped
Go 50 feet to 100 yards on level surface	Provide accessible indoor distribution vehicle; provide moving sidewalks, channelization with "short cuts"	
Level Change: Climb stairs Use escalator Use elevator Use inclined ramp	Provide each as alternative to others. Detail alternatives include inclined elevators, wheelchair elevators, stairway lift, inclined ramp, special escalators, accessible elevators	
Locate escalator	Proper use of audio, visual media to mark location	
Board ride, learn escalator	Eliminate need (provide alternatives); detail design of escalator	
Locate elevator	Broader, better use of in-terminal audio, visual media to indicate location	
Operate call button	Place button within reach of seated person; have elevator cycle automatically to call at all floors; use pressure mat on floor instead of button	
Enter elevator	Clear door opening greater than 32"; minimize floor discontinuity	
Operate elevator controls	Place buttons within reach of seated person; provide automatic operation to call at every floor	
Detect arrival at desired floor	Broader use of audio, visual media for floor information	
Exit elevator	Clear door opening greater than 32"; minimize floor discontinuity	
Ascend, descend inclined ramp	Keep ramp inclination less than 1:12; provide flat rest spots on long ramps	
Identify correct boarding point or fare barrier	Channelization (to eliminate alternatives); broader use of audio, visual media for identification of terminal area	

Library and Information Centre / Centre de documentation (SLA)
RECEIVED / REÇU LE
JUL 17 1986
Transport Canada / Transports Canada